MATHEMATICAL METHODS OF STATISTICAL QUALITY CONTROL

by K. SARKADI and I. VINCZE

AKADÉMIAI KIADÓ, BUDAPEST

MATHEMATICAL METHODS OF STATISTICAL QUALITY CONTROL

by
K. SARKADI and I. VINCZE

Quality control in industry and commercial practice is based on samples instead of a 100 per cent examination of products or lots. Consequently, correct conclusions from samples or lots can be drawn through probabilistic methods. The same is true for observed and experimental data from which general regularities are to be discovered. Although standard methods are available for these puposes, everyday life always produces new situations, necessitating the construction of additional methods, their examination and the choosing of the most economical method when several are elaborated and found suitable for the same purposes. (In this book a number of tables will help the reader in the application of the described methods.)

In respect of the above considerations, it can be said that scientific researchers, mathematicians in the field of applied sciences, engineers, specialists in quality control, as well as university students will receive great help from this book which is a standard source of verified information.

AKADÉMIAI KIADÓ
PUBLISHING HOUSE OF THE
HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

MATHEMATICAL METHODS
OF STATISTICAL QUALITY CONTROL

MATHEMATICAL METHODS OF STATISTICAL QUALITY CONTROL

By

K. SARKADI and I. VINCZE

AKADÉMIAI KIADÓ · BUDAPEST 1974

© Akadémiai Kiadó, Budapest 1974

Joint edition published by Akadémiai Kiadó, Budapest, and Academic Press,
New York and London

Printed in Hungary

CONTENTS

CHAPTER I

Introduction

1 Statistical Methods of Quality Control 11
  1.1 Inference Based on a Sample 11
  1.2 The Main Fields of Statistical Quality Control 12
  1.3 Mathematical Techniques of Statistical Quality Control 13
2 Probability Theory and Its Role in the Methods of Statistical Quality Control 14
  2.1 Random Phenomena 14
  2.2 Random Variables; The Sample Space 16
  2.3 Random Events; The Probability of an Event 19
  2.4 Utilizing the Knowledge of the Probability of an Event 22

CHAPTER II

Theoretical Foundations

1 Elements of Probability Theory 26
  1.1 Summary of Chapter I 26
  1.2 Scope of Probability Theory and Mathematical Statistics 28
  1.3 On the Classical Definition of Probability 29
  1.4 Operations with Events; Addition Theorem for Probabilities 29
  1.5 Conditional Probability 31
  1.6 Independence of Events; The Multiplication Rule for Probabilities . 33
  1.7 Independence of Experiments; Sampling With and Without Replacement 34
  1.8 Joint Examination of Two Measurements of an Item 37
  1.9 The Rule of Total Probabilities; Bayes' Theorem in the Discrete Case 40
2 Random Variables 42
  2.1 The Random Variable and Its Distribution 42
    2.1.1 The Case of One Variable 42
    2.1.2 The Random Variable in Kolmogorov's Theory 47
    2.1.3 The Case of a Vector Variable 48
  2.2 Further Properties of the Distribution Function. The Density Function of a Continuous Random Variable 49
  2.3 Characteristics of a Random Variable 54
    2.3.1 The Expected Value 55
    2.3.2 Median and Quantiles 56

2.3.3 Variance and Standard Deviation 57
2.3.4 Mean Deviation 58
2.3.5 Moments 58
2.4 Independence of Random Variables 59
2.4.1 The Notion of Independence 59
2.4.2 Correlation Coefficient 60
2.5 Distribution of Functions of Random Variables 62
2.5.1 Functions of a Single Random Variable 62
2.5.2 Sum of Independent Random Variables 63
2.5.3 Product and Quotient of Two Independent Random Variables 65
2.6 Some Properties of the Expected Value and of the Variance 66
2.7 Theorem of Chebyshev and Related Inequalities 67
2.8 Laws of Large Numbers 69
2.9 Some Important Types of Probability Distribution 70
2.9.1 Gaussian or Normal Distribution 70
2.9.2 The Poisson Distribution 74
2.9.3 Binomial Distribution 75
2.9.4 Exponential Distribution 77
2.9.5 Gamma Distribution 78
2.9.6 The Chi-Square (χ^2) Distribution 78
2.9.7 The Weibull Distribution 79
2.9.8 Student or t Distribution 80
2.9.9 The Beta Distribution 81
2.9.10 The Logarithmic Normal Distribution 82
2.9.11 The Uniform Distribution 84
2.9.12 Hypergeometric Distribution 84
2.10 The Central Limit Theorem 85
2.11 Bayes' Theorem for Continuous Variables 86
3 Fundamentals of Sampling 87
3.1 The Sample 87
3.1.1 Definition of the Sample 87
3.1.2 Sample Size 88
3.2 Sample Characteristics; the Statistic 90
3.2.1 Working Out the Statistical Data 90
3.2.2 The Most Frequently Used Sample Characteristics 94
3.2.2.1 The Sample Mean or Arithmetic Mean 94
3.2.2.2 The Order Statistics 96
3.2.2.3 The Sample Median 96
3.2.2.4 The Mean Deviation of the Sample 97
3.2.2.5 The Sample Variance and the Standard Deviation . . . 98
3.2.2.6 The Range 99
3.2.2.7 The Coefficient of Variation 100
3.2.3 The Empirical Distribution 100
3.2.3.1 The Density Histogram 100
3.2.3.2 The Empirical Distribution Function 101
3.3 Fundamentals of the Theory of Order Statistics 103
3.3.1 Distribution of the Order Statistic 103
3.3.2 Distribution of the Range 105

3.3.3 Theorems of Kolmogorov and Smirnov 107
3.3.4 Distributions of the Maximum Distances Between Two Empirical Distribution Functions 109
4 Fundamentals of Mathematical Statistics 110
4.1 Basic Problems of Mathematical Statistics 110
4.1.1 The Two Main Branches of Mathematical Statistics 110
4.1.2 Parametric and Nonparametric Problems in Mathematical Statistics 111
4.2 Fundamentals of the Theory of Estimation 113
4.2.1 The Estimation Problem 113
4.2.2 Point Estimation 114
4.2.2.1 Unbiasedness 115
4.2.2.2 Efficiency 116
4.2.2.3 Consistency 118
4.2.2.4 Sufficiency 119
4.2.2.5 The Cramér – Rao Inequality 122
4.2.2.6 Estimation of Some Important Parameters 123
4.2.3 Interval Estimation 127
4.2.3.1 The Notion of Confidence Interval. The Case of the Normal Distribution 127
4.2.3.2 Confidence Interval for the Probability 129
4.2.3.3 Confidence Interval for the Parameter of the Exponential Distribution 130
4.2.3.4 Some Important Remarks Concerning Confidence Intervals 131
4.2.4 The Method of Bayes 132
4.2.4.1 Introduction 132
4.2.4.2 The Method 134
4.2.4.3 The Lemma of Laplace 134
4.2.4.4 Two Examples of the Application of the Bayesian Method 135
4.3 Testing Statistical Hypotheses 140
4.3.1 The Statistical Test Procedure 140
4.3.1.1 Statistical Hypotheses 140
4.3.1.2 The Test Procedure 142
4.3.2 Errors of the First and Second Kind 147
4.3.3 The Power Function and the OC Curve 151
4.3.4 Uniformly Best Tests. Lemma of Neyman and Pearson 153
4.3.4.1 On Randomized Tests 158
4.3.5 Consistency and Unbiasedness 161
4.3.6 Parametric Tests of Various Hypotheses Involving Gaussian Variables 163
4.3.6.1 General Remarks 163
4.3.6.2 The u Test 164
4.3.6.3 The F Test 167
4.3.6.4 The t (Student) Test. The Welch Test 169
4.3.6.5 Comparison of the Variances of More Than Two Gaussian Random Variables, the Bartlett Test 173
4.3.6.6 Comparison of the Expected Values of More Than Two Gaussian Random Variables 174

4.3.7 Test of 'Goodness of Fit'. Tests for Homogeneity and Independence 177
    4.3.7.1 Inference Concerning the Distribution Function of a Random Variable 177
    4.3.7.2 The Test of Kolmogorov 178
    4.3.7.3 Testing for Normality 179
    4.3.7.4 Testing the Hypothesis of Exponentiality (Störmer's Test) 182
    4.3.7.5 Comparison of Two Samples. (The Kolmogorov—Smirnov Two-Sample Test) 183
    4.3.7.6 The χ^2 Test; Its Application for Testing of Goodness of Fit, of Homogeneity and of Independence 185
4.3.8 Application of Statistical Tests to the Control of Quality 189
4.3.9 Example for the Computation of Error Probabilities 191
4.4 Elements of Correlation and Regression Analysis 193
4.4.1 Estimation of the Correlation Coefficient 194
4.4.2 The Regression Curve. Regression Lines and Parabolas 197
4.5 Elements of the Theory of Decision Functions 200
4.5.1 The Decision Problem 200
4.5.2 Statistical Decision and the Decision Function 201
4.5.3 Loss Function and Risk Function 203
4.5.4 Examples of Loss and Risk Functions. The Completed Risk Function 205
4.5.5 The Role of the Unbiased Estimator and of the Sufficient Statistic 209
4.5.6 Selection of the Appropriate Decision Function. The Minimax Method 212
4.6 On Stochastic Processes 213
4.6.1 Introduction 213
4.6.2 The Concept of a Stochastic Process 215
4.6.3 The Binomial Sequence. A Gaussian Stochastic Process 219
4.6.4 The Poisson Process 223
4.6.5 The Recurrent Stochastic Process 225
4.6.6 The Pólya Process 227

CHAPTER III

Methods of Statistical Quality Control

1 Statistical Methods in the Control of Production Processes 230
1.1 General Part 230
1.1.1 Quality Control and Statistical Quality Control 230
1.1.2 Different Types of Process Control 232
1.1.3 The Outlines of Statistical Process Control 233
1.2 Process Control by Variables 233
    1.2.1. Process Analysis 233
    1.2.1.1 Techniques for the Collection of Data 234
    1.2.1.2 Evaluation of the Data and Calculation of Statistical Characteristics 237
    1.2.1.3 Determination of the Accuracy of the Machine 238

1.2.2 Adjustment of the Machine 240
1.2.3 Statistical Control During the Production Process 241
1.2.3.1 Control Charts 241
1.2.3.2 Chart for Mean and Range 243
1.2.3.3 Chart for Median and Range 249
1.2.3.4 Some Other Types of Control Charts 251
1.3 Control by Attributes 263
1.3.1 Process Analysis for Defectives 263
1.3.2 Chart for the Number of Defectives 264
1.3.3 Control Chart for Fraction Defective 267
1.3.4 Narrow-Limit Gaging 268
1.3.5 Control Charts for the Number of Defects 275
1.3.6 Other Control Charts for Attributes 277
1.4 The Choice of the Intervals Between Two Consecutive Controls . . . 278
1.4.1 Some Points of View on the Choice of the Length of the Interval 278
1.4.2 The Case of Time Intervals of Constant Length 279
1.4.3 The Case of Time Intervals of Variable Length 284
2 Acceptance Sampling 288
2.1 On Acceptance Sampling in General 288
2.2 Sampling by Attributes 289
2.2.1 Single Sampling Plan 289
2.2.2 Acceptance Sampling Schemes 293
2.2.3 Sampling Schemes Based on the Acceptable Quality Level . . . 294
2.2.4 Double and Multiple Sampling 299
2.2.4.1 Double Sampling 299
2.2.4.2 Multiple Sampling 304
2.2.5 Tightened and Reduced Inspection 305
2.2.6 Some Further Remarks Concerning Acceptance Sampling . . . 312
2.2.7 Schemes of Dodge and Romig 313
2.2.8 Some Other Approaches to Sampling Inspection by Attributes 315
2.3 Sampling Inspection by Variables 317
2.4 Sequential Sampling 330
2.4.1 On the Sequential Procedure 330
2.4.2 The Sequential Version of the Probability Ratio Test 331
2.4.3 The OC Function of the Sequential Probability Ratio Test . . . 334
2.4.4 The Expected Value of the Sample Size 337
2.4.5 Sequential Test in the Case of a Binomial Distribution 338
2.4.6 Sequential Test in the Case of a Normal Distribution 342
2.4.7 Sequential Procedure for Testing Life Span in the Case of an Exponential Distribution 346
2.4.8 Truncation and Grouping in the Sequential Test Procedure . . . 350
2.5 Acceptance Plans and the Production as a Stochastic Process 351
2.5.1 Stochastic Relation Between Sample and Lot 351
2.5.2 Sample and Lot in the Case of a Pólya Sequence 353
2.5.3 On the Acceptance Plans in the Case of a Pólya Sequence . . . 355
3 Reliability Theory 358
3.1 Failure Distribution 359
3.2 Statistical Evaluation of Empirical Failure Data 365

3.3 Problems Relating to Complex Equipment 371
3.4 Multicomponent Structures 374
3.5 Classes of Models for Failure Distributions Based on the Properties of the Hazard Rate Function 376

Appendix

Special Tables 379
Table I. Random Numbers from the Uniform Distribution 380
Table II. Random Numbers from the Standard Normal Distribution . 382
Table III. Values of the Density Function $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ of the Standard Normal Variable 384
Table IV. Distribution Function $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ of the Standard Normal Variable 385
Table V/a. Probability $p_\lambda(k)$ of the Value k of a Poisson-Distributed Variable 386
Table V/b. Cumulative Sums (Distribution) of the Poisson Probabilities 387
Table VI/a. Probability $p_n(k)$ of the Value k in the Case of the (n, p) Binomial Distribution 388
Table VI/b. Distribution Function of the Binomial Law 390
Table VII. Table for the F Test at One-(Two-)Sided Levels 95% (90%), 99% (98%), 99.9% (99.8%) 392
Table VIII. Critical Values of the t (Student) test at One-(Two-)Sided Levels 97.5% (95.0%), 99.5% (99.0%) and 99.95% (99.9%) 397
Table IX. Quantiles of the χ^2-Variables, i.e., Critical Values of the χ^2 Test 398
Table X. Table to Kolmogorov's One Sample Test (Two-Sided Levels 95% and 99%) 399
Table XI. Critical Values to the Kolmogorov—Smirnov Two Sample Test ($m = n$, Two Sided Levels 95% and 99%) 399
Table XII. Critical Values of the Wilcoxon Two Sample Test at One-(Two-) Sided Level 97.5% (95%) 400

Bibliography 401
Textbooks 401
Articles 403
Tables 407
Standards 408

Index 409

CHAPTER I

INTRODUCTION

1 STATISTICAL METHODS OF QUALITY CONTROL

1.1 INFERENCE BASED ON A SAMPLE

Research and practice in the fields of industry, agriculture, and commerce have for many years utilized sampling procedures in evaluating product quality. In the case of a mass product, such as screws or nails, it is generally uneconomical to examine each item or every detail. In many instances, moreover, destructive testing procedures must be used in order to determine the quality of a product—breaking tests on metal parts, lifetime tests for light bulbs and telecommunications parts, etc. Obviously, testing of each item is out of the question in such instances.

Evaluating quality on the basis of a sample is not as straightforward as it might seem, however. Suppose we wish to know the quality of a lot of 10 000 screws. If, instead of examining every item, we examine only a sample—say 200 screws—we can draw no conclusion about the entire lot with certainty. Chance will determine which pieces are included in the sample. Therefore our inference may be wrong: a sample from a lot of good quality may contain a large number of defectives; a sample from a lot of poor quality may contain relatively few defectives—perhaps even none.

For example, a sample containing 200 screws might include only 4 defective items (2% of the sample); yet this sample might come from a lot containing 8% or more defectives. On the other hand, a sample containing 16 defectives (8% of the sample) might be drawn from a lot containing only 2% defectives. Thus the result of a sampling procedure depends upon chance.

In order to draw correct conclusions from random samples we need to make use of probability theory, which concerns itself with the laws of random phenomena. In quality control a branch of probability theory known as mathematical statistics is employed: hence the term statistical quality control. *Statistical methods of quality control are those sampling procedures that apply probability theory and mathematical statistics.*

As already stated, a single random sample does not enable us to draw a conclusion with certainty; we must always be aware of the possibility that a decision made on the basis of the sample may be wrong. Mathematical statistics enables us, however, to obtain a quantitative measure of the reliability of a sampling procedure. We can construct statistical sampling methods having a known average ratio of wrong decisions: in a long run of samples, such a ratio predicts the fraction of wrong decisions that will be made. Furthermore, sampling procedures with higher or lower ratios of wrong decisions can be constructed. A low ratio ensures greater safety but sometimes requires an extremely large sample. Hence a low level of wrong decisions and a low sampling cost (i.e., a small sample) are contradictory requirements. When local circumstances and technical conditions are taken into account, mathematical statistics helps us to find the most suitable sampling procedures in any given instance.

1.2 THE MAIN FIELDS OF STATISTICAL QUALITY CONTROL

There are two main forms of statistical quality control: *acceptance sampling*, which deals with finished or semifinished goods; and *control during processing*. In both forms, technical or economic limitations in most cases make 100% examination of the product impractical. Therefore a control system based on sampling and mathematical statistics is required.

Acceptance sampling can always be applied to the examination of a lot of finished or semifinished goods. Such sampling reveals the presence of defectives, and the continuous application of such methods provides much information that can be utilized in improving the production process—though of course such improvements affect only future production.

Statistical methods of control during processing have a smaller field of application: to be considered a random process the production must be essentially independent of external intervention for long periods. Nevertheless, statistical methods are of great significance when applicable. In applying statistical control during processing we obtain information about the trends of the process; *such information may direct our attention to the danger of defective production in advance, while the size or quality of the examined pieces is still good.*

Both forms of statistical quality control may be divided into two categories according to the criteria used in evaluating the samples. In *control by variable* samples are evaluated on the basis of measurements of appropriate geometrical, physical, or chemical properties of the product. In *control by attribute* only the presence or absence of certain attributes—for example, whether the item is good or bad—is considered. (We shall not be concerned with multiple classifications of attributes, such as, for example, good, repairable, and defective.)

So far we have spoken in terms of products that consist of separate pieces, not about "continuous" products such as occur, for example, in the textile, metallurgical, and chemical industries. This book considers mainly methods elaborated for products consisting of pieces, many of which methods can be applied to continuous products if the products are divided into units in some way—textile goods into units of 1 meter, 3 meters, etc. However, some direct methods for continuous products will also be described.

1.3 MATHEMATICAL TECHNIQUES OF STATISTICAL QUALITY CONTROL

As already mentioned, any well-founded control system based on samples has to apply the methods and results of probability theory and mathematical statistics. We meet quantities subject to the influence of chance not only in production processes but in most aspects of industry concerned with economic and organizational problems or business decisions. Consequently, it is reasonable to apply probabilistic methods in nearly all branches of industrial management.

Once a statistical quality control program has been instituted, one has only to be acquainted with some arithmetic operations in order to carry out the technical calculations involved. However, the introduction and continuous direction and supervision of such methods require the knowledge of the theoretical background as well.

The aim of this book is to introduce to the reader the part of the theory of probability and mathematical statistics closely connected with industrial applications and to acquaint him with the most important statistical methods of quality assurance. With this knowledge it is possible to select an appropriate method in any given situation, to combine different methods, and to construct new ones if needed.

Because probability theory and mathematical statistics are mathematical disciplines, knowledge of calculus (the theory of elementary functions, the theory of differentiation and integration), combinatorics, and the elements of algebra and geometry is a prerequisite for understanding this book.

In some writings on the theory of probability—even some of those that are closely connected with applications—an esoteric formalism is used. Our aim is to treat the subject as simply as possible. Extremely abstract formulation will be avoided, as will mathematical details (for example, proofs will generally not be given). However, the main concepts that will help the reader to understand the statistical literature will be outlined in the general part of the book.

While striving for simplicity, we want to maintain mathematical precision. An exact mathematical foundation of the theory of probability was given by Kolmogorov in 1933; our approach may be regarded as a simplified form of that used by Fraser [A 17]. Rigorous treatment and great development of the theory led to a wide extension of the field and of its applications. Furthermore, it helped to avoid inadequate applications in practice.

In order to consider statistically a random process taking place in reality it is necessary to relate certain theoretical concepts to the process. For example, the real process corresponds to a random experiment, the quantity that is subject to error to a random variable, the set of different possible outcomes of the experiment to a sample space, and so on. In the absence of a well-defined statistical model there is no possible way of knowing whether a certain event does or does not have a probability at all. An event may be a random event having a definite probability in one model, but not in another one. On the other hand, a model constructed for a given problem might be extended if further questions arise.

The remainder of this chapter is devoted to a rather detailed explanation of essential concepts. In Chapter II a concise treatment of the fundamentals of probability theory will be given. Chapter III throws light on the connection between theoretical and empirical quantities and the importance of the theoretical concepts in applications.

2 PROBABILITY THEORY AND ITS ROLE IN THE METHODS OF STATISTICAL QUALITY CONTROL

2.1 RANDOM PHENOMENA

Mathematical statistics is a discipline distinct from the use of descriptive statistics such as economic, political, or demographic statistics-although it can be applied in these fields as well as in numerous others. Mathematical statistics can be applied in all situations that require the use of probability theory. Such situations will be examined after we give the following definition of probability theory:

> *Probability theory deals with the discovery and mathematical description of the objective regularities occurring in random phenomena or in random experiments.*

Here we speak of a *phenomenon* as something brought about by nature—e.g., the weather or mean temperature on a given day of the year, the nature of the soil in a given agricultural area, the water content of the air, etc. We speak of an *experiment*, in general, as the result of human activity. In this sense not only are there scientific and industrial experiments but, for example, the production process of a given machine under given circumstances is also considered to be an experiment. So long as the essential circumstances do not change, the production of consecutive pieces will be regarded as the same experiment repeated many times. When speaking generally about objects of probability theory we use *phenomenon* and *experiment* (and sometimes *trial*) interchangeably.

We shall consider phenomena or experiments that occur or may be carried out under given, known conditions. Even if the essential circumstances under which an experiment takes place are well defined and controlled, we may have a *random* experiment (phenomenon) if—as in reality almost always occurs—a large number of small effects influence the final result. In the repeated observation of a phenomenon or the repeated performance of an experiment under essentially unchanged conditions, the outcome will differ to a greater or lesser extent from case to case.

Now let us consider two basic examples:

Experiment 1. Suppose we have a lot of cogwheels consisting of N (e.g., 5000) items. The experiment is to select randomly n ($< N$) pieces from the lot. The random selection can be carried out with the aid of a table of random numbers (see Table I, on p. 380).

Let us suppose that we are able to state whether each of the sample pieces is good or defective. If the lot contains M ($> n$) defective pieces and $N - M$ good ones, then the random samples may contain any of the following possible numbers of defectives: $0, 1, 2, \ldots, n$. Repetition of the experiment will show that the resulting number of defectives changes from one trial to another. (Before repeating the experiment we return the preceding sample to the lot.)

Experiment 2. An automatic machine is set to produce cylinders having prescribed dimensions.

Even if we ensure uniform working of the machine, we will find that the different dimensions fluctuate from piece to piece. This fluctuation is due to the fact that the dimensions of different pieces will be influcenced by a great number of small, unaccountable effects, such as vibration of the machine, small fluctuations in temperature, inhomogeneity of material, etc.

Let us assume, for example, that a nominal value of 23 mm is prescribed for the diameter, and that values smaller than 22 or larger than 24 mm have nct been found to occur in our experience.

In both of the preceding examples we have *random experiments*: The essential conditions are fixed, but the outcome is not known in advance; the data resulting from repeated trials will show fluctuations.

Random phenomena must satisfy a further condition if the theory of probability is to be applied to them: they must be *mass phenomena*. That is, they must be repeated many times, or at least capable of such repetition. This is the case in both experiments described above. Thus we may state: *The theory of probability considers random, mass phenomena.*

It is of interest to note the "two different kinds" of randomness in the two experiments. In Experiment 2 randomness is a consequence of the experiment itself, which is the result of a complicated process that cannot be controlled in its smallest details. On the other hand, in Experiment 1 randomness is introduced by using random numbers to select the items in each sample; this "randomization" enables us to apply probabilistic methods.

2.2 RANDOM VARIABLES; THE SAMPLE SPACE

Considering Experiment 2 we may observe that a precise description of the "outcome" or "result" of an experiment (if such description is possible at all) may require a very large number of characteristics—geometric, physical, chemical, and other technical data. But for practical reasons, we limit our analysis of the experiment to one or more of these quantities.

The outcome or result of an experiment is in this case the observed value (or values) of the considered quantity (or quantities). Before carrying out a (random) experiment we cannot predict the result precisely. We can, however, state the possible results that may occur. In a random experiment the quantities we are interested in are random quantities, called *random variables*. Experimentation (i.e., an observation on a random variable) yields a determined value, which we call an *observed value* of the random variable. Considering the first experiment, let us denote by $\varkappa$ the number of defective items in a sample of size n taken randomly from the lot. Carrying out the experiment we obtain a certain integer k as the observed value of the random variable $\varkappa$.

Although the value of the random variable $\varkappa$ is unknown before experimentation, we do know that the possible outcomes are $0, 1, 2, \ldots, n$. The set of all possible values that a random variable (or variables) can assume is called the *sample space* for the experiment, denoted in the following by Gothic letters $(\mathfrak{X}, \mathfrak{Y}), \mathfrak{Z}, \mathfrak{K}, \ldots)$.

In Experiment 1, considering the random variable $\varkappa$, the sample space is

$$\mathfrak{K} = \{0, 1, 2, \ldots, n\}.$$

Note that this variable (the number of defective items in the sample) can have only integer values; in other words it is a *discrete variable*.

In Experiment 2, one random variable is the diameter of the cylinder produced by the machine. Denoting this random quantity by X, the possible values of X belong to the interval (22.0, 24.0), as already mentioned. However, an interval larger than the expected fluctuation is often chosen as the sample space. Depending on the type of fluctuation, this interval may be $(0, \infty)$ or even the whole real axis. Let us choose in our experiment the following sample space for X:

$$\mathfrak{X} = \{-\infty < x < \infty\}.$$

This choice simplifies the mathematical description of the phenomenon, although in practice negative diameters obviously never occur.

In the same experiment suppose that besides the diameter a further quantity is of practical importance—namely, the breaking strength. Let us denote this random variable by Y.

We now have a pair of random variables (X, Y), which form a *random vector*. If in an observation we find the value $X = x$ for the diameter of a piece, and $Y = y$ for the breaking strength of the *same* piece, then we say that the observed value of the vector variable is (x, y). The sample space in this case is a two-dimensional domain—the Cartesian plane or a part (subset) of it. This is a natural extension of the case of a single random variable. A suitable choice of the sample space for this case is

$$\mathfrak{Z} = \{(x, y)\colon\ -\infty < x < \infty;\ 0 < y < \infty\},$$

which is the upper half of the (x, y) plane.

Note that in this experiment the variables (physical measurements) can take on any value within the limits of the sample space—not just integer values. Such variables are called *continuous variables*.

Let us now turn to a third experiment:

Experiment 3. In a spinning mill we examine the work of a given spinning machine. Let us consider the random variable ν, the number of yarn breakages during a shift. The value of ν on any trial will be an integer k. Even in the case of a good machine and a homogeneous quality of yarn, this number will differ from day to day. Its possible values are 0, 1, 2, . . .—i.e., we do not restrict the number of possibilities, in spite of the fact that in reality the breakages will never exceed a finite number limited by the technical circumstances. Accordingly, the sample space is

$$\mathfrak{N} = \{0, 1, 2, \ldots\}.$$

At the same time let us consider as a further random variable the quantity Z of yarn produced by the machine during a shift. To consider these two variables simultaneously, we would use a random vector (ν, Z) having the sample space

$$\mathfrak{Y} = \{(k, z)\colon\ k = 0, 1, 2 \ldots;\ 0 < z < \infty\}.$$

In this example the vector consists of one discrete and one continuous variable.

Many experiments have a qualitative, rather than quantitative, outcome. The result of an experiment may be, for example, the presence or absence of a certain property, the color of a product, etc. Even in such cases we may assign numbers to the different possibilities. Let us suppose, for example, that in Experiment 1 we draw from the lot a single piece, which may be defective or good. A new random variable X can be defined in the following way:

$$X = \begin{cases} 1 & \text{if the item is defective,} \\ 0 & \text{if the item is good.} \end{cases}$$

This definition is in accordance with the definition of $\varkappa$ where $n = 1$, as X is now the number of defectives in a sample of size 1. On the other hand, selecting n pieces we may take into account the succession of the pieces and define the following random variables:

$$X_i = \begin{cases} 1 & \text{if the piece selected the } i\text{th time is defective,} \\ 0 & \text{if the piece selected the } i\text{th time is good.} \end{cases}$$

In this case the random variable considered is an n-tuple $(X_1, X_2, \ldots, X_n)$ —i.e., a random vector —and the sample space contains all n-tuples consisting of 0's and 1's:

$$\mathfrak{E} = \{(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \colon \varepsilon_i = 0 \text{ or } 1;\ i = 1, 2, \ldots, n\}.$$

The number of different elements in $\mathfrak{E}$—i.e., the number of different possibilities of the outcome—is 2^n.

As can be seen easily, the following simple relation holds between the random variable $\varkappa$ and the X_i's:

$$\varkappa = X_1 + X_2 + \ldots + X_n.$$

Summarizing: In the application of mathematical statistics these steps should be followed: (1) specify the experiment, (2) ensure essentially unchanged conditions of observation (i.e., the possibility of repeating the experiment), (3) define the random variable or variables of interest, and (4) fix the sample space of the random variable or random vector.

"Infinite" sample spaces and "continuous" quantities are mathematical abstractions, which enable us in most cases to construct adequate and easily manageable mathematical models of the phenomena.

2.3 RANDOM EVENTS; THE PROBABILITY OF AN EVENT

In Experiment 2 assume that the machine produces 100 cylinders. Measuring the diameter of each piece, we count the number of pieces whose diameter does not exceed a certain value—say, 22.88 mm. Suppose we count four such pieces in the series.

The preceding can be expressed as follows: As regards Experiment 2 we have selected an event—the occurrence of a piece whose diameter X is no larger than 22.88 mm. We shall use the notation $\{X \leqslant 22.88\}$ for this event. In the series of 100 trials the event occurred four times; in other words, the *frequency* of this event in the sequence of 100 trials was found to be 4.

An event is a part or a subset of the sample space defined for an experiment. Once an event has been specified, it will sometimes be denoted by a capital letter—e.g., $A = \{X \leqslant 22.88\}$. Capitals will also be used when speaking about an event in general.

If an event B does not occur in a trial, then the complementary set occurs. We call the corresponding event the *complementary event* $\bar{B}$. For example, in the case under discussion we have $\bar{A} = \{X > 22.88\}$. The frequency of $\bar{A}$ in the sequence of 100 trials was $100 - 4 = 96$.

If in a sequence of trials of length n a given event B occurs exactly k times, then we say that the *frequency* of B in the given sequence of trials is k; the *relative frequency* of the event B in the same sequence of trials is k/n. For example, in the cylinder experiment the relative frequency of event A was $k/n = 4/100 = 0.04$; the relative frequency of $\bar{A}$ in the same sequence of trials was $(n - k)/n = 0.96$.

We can make general statements about a sequence of trials of length n without actually performing the experiment. In this case the frequency of an event B is a random variable $\varkappa_n$; its value may be observed by carrying out a sequence of trials of length n. The observed value of $\varkappa_n$ will vary in different sequences; so will the relative frequency $\varkappa_n/n$—i.e., relative frequency is a random variable too.

Now the crucial experience that can be gained from many series of observations is this: When a random experiment based on observing a given event is carried out repeatedly, we shall observe that the relative frequency of the event will fluctuate from experiment to experiment. This fluctuation will take place *around a definite value*. Generally, the greater the length n of the series, the smaller will be these fluctuations. In other words, we may find large deviations among the relative frequencies, but this happens very rarely if the number n is very large. Anyone may convince himself of this fact by making trials in large numbers, e.g., throwing dice, selecting cards, or observing several random quantities. The theoretical explanation of this experience is contained in the law of large numbers (II, Section 2.8).

The relative frequency of an event in a sequence of trials is an *empirical* quantity. The *theoretical* value around which the relative frequency fluctuates is called the *probability* of that event. (Figures 1 and 2.) The probability of an event is an objective measure that can be determined by means of long sequences of observations. In certain cases it is also possible to determine the probability of an event theoretically; but even in such cases the value obtained theoretically has to be controlled by experience. This requires statistical methods. As a further illustration one may carry out Experiment 2 for a series of length $n = 100$ many times, and obtain, for example, the following values for the relative frequencies of the mentioned event:

$$\frac{3}{100}, \frac{2}{100}, \frac{4}{100}, \frac{2}{100}, \frac{1}{100}, \frac{3}{100}, \frac{3}{100}, \frac{4}{100}, \frac{2}{100}, \frac{1}{100}, \frac{4}{100}, \frac{0}{100}.$$

These fractions vary between 0.00 and 0.04. Now taking $n = 1000$, we may find the following values for the relative frequencies:

$$\frac{32}{1000}, \frac{21}{1000}, \frac{37}{1000}, \frac{34}{1000}, \frac{19}{1000}, \frac{29}{1000}, \frac{26}{1000}, \frac{38}{1000}, \frac{22}{1000}, \frac{25}{1000}.$$

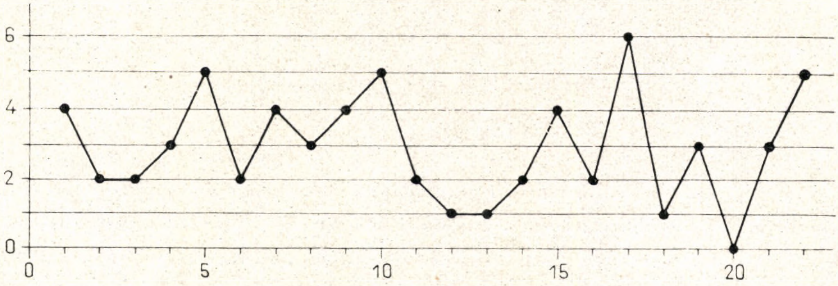


FIG. 1. Fluctuations of relative frequencies in repeated sequences of trials (sequences of length $n = 6$; coin tossing with a homogeneous coin)

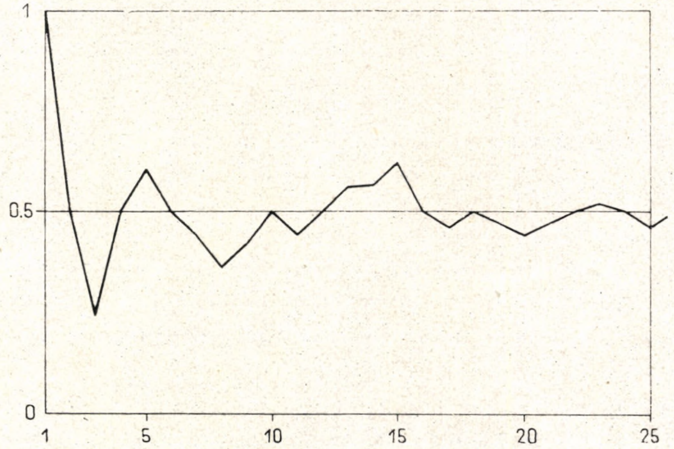


FIG. 2. Fluctuations of relative frequencies for different lengths n of sequences (coin tossing with a homogeneous coin)

These values are closer to each other (and thus to a certain fixed value) than are the relative frequencies for $n = 100$.

The probability of an event A is denoted by $P(A)$. For example, in the experiment under discussion,

$$P(A) = P(X \leqslant 22.88).$$

$P(A)$ is the value around which the relative frequencies k/n will fluctuate in subsequent sequences of trials or in the same sequence of trials when n is increased. Because the relative frequency of $\bar{A}$ is $(n - k)/n$, the value around which these fractions will vary is $1 - P(A)$; that is,

$$P(\bar{A}) = 1 - P(A).$$

Furthermore, because $0 \leqslant k/n \leqslant 1$, the probability for any possible event A also falls within these limits:

$$0 \leqslant P(A) \leqslant 1.$$

In a practical experiment such as a manufacturing process, conditions sooner or later alter. This fact, however, does not contradict the validity of the concept of probability. For example, if the tool in Experiment 2 becomes worn after a certain time, the relative frequencies still fluctuate, but now around a different value; in other words, the probability of the event $\{X \leqslant 22.88\}$ has changed. Because the essential conditions have altered, obviously the experiment is different from the original one. When the change of probabilities of events with time is taken into account, we have what is called a random or stochastic process.

2.4 UTILIZING THE KNOWLEDGE OF THE PROBABILITY OF AN EVENT

Let us assume that in the production process in Experiment 2, the event $\{X \geqslant 22.88\}$ represents the required standard for the diameter of an item. In this case the occurrence of the event $A = \{X < 22.88\}$ means that the item is defective. Let the probability of this event be $P(A) = 0.033$.

Suppose now that the conditions of production remain unchanged. Examining subsequent series of 100 pieces each we will find that the number of defective items fluctuates around 3 or 4 pieces. In series containing 1000 items each, a fluctuation will take place around 33. Although we cannot tell in advance how many defective items will be found in a given series of 1000 pieces, we know that this number will not differ greatly from 33. The larger the sequences, the smaller the deviation of the relative frequencies from the average value.

Knowing the probability of the event, information about the outcome of the experiment on future trials may be inferred, provided the conditions do not change. Even if we cannot foresee the exact outcome of the next experiment, we may survey the phenomenon relatively well in its entirety.

In the practice of production, however, conditions cannot be expected to remain unchanged for long periods: sooner or later, some change will take place. For example, an improvement of the technology may result in a lower value of the percentage around which the relative frequency of the event $\{X \leqslant 22.88\}$ fluctuates. On the other hand, if some irregularity occurs (e.g., the cutting tool becomes worn or slackens), we may get fluctuations of the relative frequency around a larger value than before.

How can the theory of probability be utilized in controlling the production process? As we shall see later, statistical quality control begins with a *preliminary sampling*. The purpose of this procedure, in our example, is to determine the probability of the occurrence of defective items. Let us suppose that the value $0.033 = 3.3\%$ was obtained by means of a preliminary sampling. Controlling the production process now involves testing the *hypothesis* that the production process is stable—i.e., that the probability of the event $\{X \leqslant 22.88\}$ is constant. If, for example, half an hour after the preliminary sampling, we check ten items and find two of them defective, we draw the conclusion that the process is not stable; the machine has to be supervised. If our hypothesis that $P(X \leqslant 22.88) = 0.033$ holds, then among 100 items the number of defectives will be in most cases between the limits 2 and 6. If two defectives are found in a sample of only ten items, this seems to contradict our hypothesis.

Let us emphasize once more, in connection with random phenomena, that it may happen by chance that two defectives will be found among ten items, even if the probability of defective items is only 0.033. But, as we shall see later, such an event is very rare. If from such a sample we draw the conclusion that changes in the production process are necessary and therefore undertake corrective measures, this decision will turn out to be economical in the long run.

Let us consider now the case described in Experiment 1. There we had a lot of cogwheels containing $N = 5000$ pieces. In practical situations we do not know the actual value of the fraction defective, but let us suppose that a value of 3% or less is tolerable. We therefore assume that the number of defectives in an acceptable lot is $M = 150$ (i.e., 3% of the lot) and wish to test this hypothesis. If we select a sample of size $n = 120$ from the lot randomly, the number $\varkappa$ of defectives in the sample is a random variable, the set of its possible values being $\{0, 1, 2, \ldots, 120\}$.

The quality control procedure is based on a sample; intuitively it is obvious that if the observed value of $\varkappa$ is large, the lot should be regarded as bad and should be rejected. If, however, the observed value of $\varkappa$ is small, we draw the conclusion that the lot probably has a fraction defective not exceeding $0.03 = 3\%$, and therefore we accept it. But what is to be regarded as "large" and what as "small" for $\varkappa$?

The concept of probability can be utilized to test whether a given limit on $\varkappa$ is suitable for practical use. Let us arbitrarily choose 6 as a limit for still not "large" values of $\varkappa$, and let us denote by C the event that $\varkappa$ is smaller than or equal to 6:

$$C = \{\varkappa \leqslant 6\} = \{\varkappa = 0, 1, 2, \ldots, 6\}.$$

Now we determine the probability of this event in the context of Experiment 1, supposing the number of defectives to be $M = 150$ (3%). In a large number, say m, of experiments, we count the number k of cases in which the outcome is C—i.e., where the number of defectives in the sample of size $n = 120$ does not exceed 6. We treat the relative frequency k/m (m being large enough) as a value that lies near to the probability

$$P(C) = P(\varkappa \leqslant 6).$$

This procedure is troublesome, but, as we shall see later, the probability $P(C)$ can be determined very simply in a theoretical way. Let us assume that we obtain the value $P(C) = 0.99$. This means that if all the lots are good, then repeated sampling will yield 6 or fewer defectives in 99% of the cases, and 7 or more defectives in only 1% of the cases.

The foregoing test indicates that 6 can indeed be considered a "small" value for the upper limit on $\varkappa$. In checking future lots, therefore, if we find 7 or more defective items in a sample of size $n = 120$ items, we will reject the hypothesis that the lot is good: we will not accept the lot. If sampling yields 6 or fewer defective items, the lot will be accepted.

As pointed out already in Section 1.1, either inference may be wrong: (a) It might happen that a sample from a good lot contains, by chance, more than 6 defective items. In this case we reject a good lot. This is called an *error of the first kind*. In our example, such errors will occur on 1% of all good lots tested. (b) It may happen that a lot containing, for example, 500 defective items (10% of the lot) will yield a sample containing 6 or fewer defective items, and so we accept a bad lot. Such a mistake is an *error of the second kind*. We could of course construct a procedure to determine the probability of this kind of error too.

Mathematical statistics helps us to diminish the occurrence of faulty inferences, or at least to make our protection against such errors economically optimal.

We would like to emphasize the following: (1) Inferences based on a random experiment do not necessarily lead to true statements. Therefore we do not say, for example, that a given lot is good or that it is bad. A statistical procedure leads to a *decision*, which may be a correct decision or a wrong one. (2) A statistical procedure is adapted to and valid for mass phenomena. In single cases we may make wrong decisions, but if we use the method over a long period of time, most of our decisions will be correct and the procedure will be economical. (3) It is possible to construct statistical procedures for a given problem that have very small probabilities for errors of both the first and the second kind. But achieving small probabilities of error requires in general very large samples. Such sampling is reasonable only in exceptional cases—e.g., when good quality is of special interest (medicine, etc.) or when the procedure is to be used in only a small number of cases.

The examples in this chapter have illustrated how we proceed from experience to the construction of theoretical concepts and how these concepts are utilized in practice. The aim of these considerations has been to clarify the fundamental and essential point of statistical methods in quality control. The engineer or economist who wishes to apply statistical methods in his special field has to acquire a particular special way of thinking, the mentality and logic of probability and especially of statistical inference. A more thorough explanation of the basic concepts already mentioned will be given in Chapter II.

CHAPTER II

THEORETICAL FOUNDATIONS

1 ELEMENTS OF PROBABILITY THEORY

1.1 SUMMARY OF CHAPTER I

We begin this chapter by summarizing the concepts introduced in Chapter I, adding some new definitions as we proceed.

Probability theory deals with regularities appearing in random mass phenomena (i.e., problems in which random elements occur). The first task in dealing with such a problem is to specify the phenomenon or experiment, to set up the main conditions under which it occurs. Having fixed the experiment, we select, according to our practical interest, one or more quantities, which are to be regarded as random variables. In cases where we are interested in some property or quality that has no size or measure, we assign different numbers to the different possibilities—e.g., in throwing dice, we assign the integers 1 through 6 to the six sides of each die.

In repeated trials the observed values of the random variables are subject to random fluctuations. The set of all possible values that the random variables may take on is called the sample space for the experiment.

An event is a certain part (subset) of the sample space, consisting of certain possible outcomes of the experiment. The outcomes forming an event will be specified by the practical needs of the problem. For example, in connection with Experiment 1, we specified the event $C = \{\varkappa \leqslant 6\}$ —i.e., the event that a given sample contains 6 or fewer defective items.

In the case of discrete random variables, all possible subsets may be considered as events. In the case of continuous random variables there are some restrictions imposed on the choice of subsets by the measure-theoretic technicalities of probability theory. But these restrictions cause no practical difficulty, since the subsets of interest are generally intervals, half-lines and combinations of these domains which are all admissible events. When two, three, or more random variables are involved, the sample space has the corresponding dimensions; thus with two variables the subset can be a part of a plane; with three a region in the three-dimensional space; and so on.

A single possible outcome of an experiment—that is, the observed values of all random variables on one trial—is called an *elementary event* or a *point in the sample space.* Accordingly, an event is a set of elementary events, or a set of points in the sample space.

If, in a given experiment, we have defined an event A, the points of the sample space that do not belong to event A form the *complementary* event $\bar{A}$.

The event that contains all possible outcomes of an experiment—i.e., the whole sample space—is called the *certain event:* its notation will be I throughout. Whenever the experiment is carried out, the event I is certain to happen. The complementary event $\bar{I}$ is the *impossible event* or the empty set $\emptyset$.

If we carry out an experiment n times, out of which the event A occurs k times (and thus the event $\bar{A}$ $(n - k)$ times), then k/n is the relative frequency of event A in this sequence of trials. The relative frequency is subject to random fluctuations and is therefore a random variable. In successive sequences of trials, these fluctuations take place around a well-defined value, the probability of A, written $P(A)$. Experience shows that the longer the sequences of experiments, the smaller the fluctuations of the relative frequency; in long sequences of trials, the occurrence of large deviations is very rare.

Because the relation

$$0 \leqslant k/n \leqslant 1$$

holds for the relative frequency, the corresponding relation

$$0 \leqslant P(A) \leqslant 1$$

must be valid for the probability.

For the probability of the complementary event, the following holds:

$$P(\bar{A}) = 1 - P(A), \qquad P(A) + P(\bar{A}) = 1.$$

Finally, it is obvious that for the certain event I and for the impossible event $\emptyset$, respectively,

$$P(I) = 1, \qquad P(\emptyset) = 0.$$

We wish to remark that $P(A)$ is the probability of A in a *single* experiment. However, probability can be interpreted most easily by means of sequences of trials, as we have done in the preceding paragraphs.

1.2 SCOPE OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS

Suppose we have an experiment and an event corresponding to this experiment—e.g., Experiment 2 and the event that the diameter of the cylinder is less than 22.88 mm. This event has been denoted by A. We assume that its probability is known: $P(A) = 0.033$.

In carrying out a single experiment, either the event A or event $\bar{A}$ occurs.

In order to clarify the main features of probability theory—even if in a rather simplified form—we raise the following question: Knowing $P(A)$, what is the probability of the event $\{\varkappa = k\}$ $(k = 0, 1, 2, \ldots, n)$, where $\varkappa$ is again the number of experiments in which the event A has occurred. According to our present knowledge we can determine, or rather approximate, such probabilities through experience gained from a sequence of experiments containing n trials each.

However, the probability can be derived theoretically.

A fundamental task of the theory of probability is to determine the probabilities and other theoretical quantities of complicated events when the probabilities of certain simple events are known. Thus, from the known probabilities we may draw conclusions concerning the future behavior of complex random quantities.

Mathematical statistics deals, in a sense, with problems of the opposite character: How to make inferences, respecting theoretical quantities, from observed values. Such inferences usually belong to one of the following types: (1) to determine a quantity from observed data; and (2) decide on the basis of observed data whether two theoretical quantities can be considered equal or not. The first of these belongs to the branch of mathematical statistics called the *theory of estimation*. If the probability $P(A)$ were known, for example, and the event A was observed 5 times in 100 independent trials, what is the "best" estimate of $P(A)$? How good is that estimate? The second fundamental type of problem in mathematical statistics belongs to the *theory of testing hypotheses*. For example, suppose we have two lots of cylinders produced by two different machines. Can these lots be mixed up and regarded as a homogeneous lot? This must be decided by taking samples from both lots.

The development of the theory of mathematical statistics is due to R. A. Fisher, J. Neyman, and A. Wald. Wald has given a unified treatment of the two main problems above in the *theory of decision functions*.

1.3 ON THE CLASSICAL DEFINITION OF PROBABILITY

We have defined the probability of an event as the value around which the relative frequencies obtained in sequences of trials fluctuate. This definition—justified in practice by much experience—can be justified theoretically by the law of large numbers. We shall turn to this theorem later.

Let us now compare the definition just given with the classical definition of probability. According to the classical definition—but using modern terminology—the probability of an event is a fraction whose denominator is the number of all elementary events and the numerator is the number of elementary events contained in our event. This definition, however, is restricted to finite sample spaces, and it assumes implicitly that all elementary events have the same probability.

According to the classical definition, the probability of the event of getting a 5 in a toss of a die is $\frac{1}{6}$. This may be true if the die is a perfect cube and if its material is homogeneous. If, however, these conditions are not fulfilled, the probability may be some other fraction.

Our definition, on the other hand, satisfies the theoretical requirements and is in agreement with experience as well. In the example just given, for instance, the statement that the probability equals $\frac{1}{6}$ can be regarded as a theoretical approach only; it can be justified by proving that it agrees with experience.

1.4 OPERATIONS WITH EVENTS; ADDITION THEOREM FOR PROBABILITIES

In connection with an experiment we say that two events are *disjoint* if they have no elementary event in common. This means that if in a trial of the experiment one of them occurs, then the other cannot take place. For instance, in Example 3 of Chapter I, $A = \{\nu \leqslant 16\}$ and $B = \{16 < \nu \leqslant 36\}$ are disjoint events.

The event that contains all elementary events of A and of B is called the *sum* of the two events; in the above example

$$A + B = \{\nu \leqslant 36\}.$$

The event $A + B$ occurs in a trial if either A or B occurs.

The sum of two events has the same definition if the events are not disjoint. Let

$$A' = \{\nu \leqslant 20\} \qquad \text{and} \qquad B' = \{10 \leqslant \nu \leqslant 40\}$$

then $A' + B'$ is a set of all elementary events contained in A' or in B':

$$A' + B' = \{\nu \leqslant 40\}.$$

If the two events are not disjoint, then their sum means that either of the events—or both of them—occur.

The events A' and B' have a common part. This we call the *product* of the two events:

$$A'B' = \{10 \leqslant \nu \leqslant 20\}.$$

The event $A'B'$ occurs if the events A' and B' occur at the same time.

If A and B are disjoint, then the probability of event $A + B$ equals the sum of the probabilities of the events A and B:

$$P(A + B) = P(A) + P(B). \tag{1}$$

This is an immediate consequence of our definition of probability. Let us conduct a sequence of trials of size n; denote the number of trials by k_A and k_B when A and B, respectively, have occurred. Then in the same sequence of trials the frequency of the event $A + B$ is $k_{A+B} = k_A + k_B$. But $k_{A+B}/n = (k_A/n) + (k_B/n)$ fluctuates around $P(A + B)$, while k_A/n and k_B/n fluctuate around $P(A)$ and $P(B)$, respectively. It is also true that

$$P(A + B + C) = P(A) + P(B) + P(C)$$

if A, B, and C are pairwise disjoint. Turning to the general case and denoting the frequencies of the events in a sequence of trials by $k_{A'}$, $k_{B'}$, $k_{A'B'}$, and $k_{A'+B'}$, it is easy to see that the following relation holds:

$$k_{A'+B'} = k_{A'} + k_{B'} - k_{A'B'}.$$

The same holds also for the relative frequencies and as a consequence for the probabilities around which they fluctuate:

$$P(A' + B') = P(A') + P(B') - P(A'B'). \tag{1$'$}$$

This is the addition theorem for probabilities, valid for any pair of events in the case of an experiment.

If two events A and B are disjoint, then their product AB is the empty set $\emptyset$, with probability 0; i.e., $P(AB) = P(\emptyset) = 0$, and Eq. (1′) takes on the form of Eq. (1).

Similar considerations lead to the relation

$$P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC).$$

Turning to Example 2, let us assume that a cylinder is good if its diameter X lies between 22.88 and 23.12 mm. In this case the event D that an item is defective is

$$D = \{X \leqslant 22.88\} + \{X \geqslant 23.12\}.$$

Because the two events on the right-hand side are disjoint, the probability of an item being defective is

$$P(D) = P(X \leqslant 22.88) + P(X \geqslant 23.12).$$

If the events $A_1, A_2, \ldots, A_r$ have the following properties—(a) they are pairwise disjoint:

$$A_iA_j = \emptyset; \quad i \neq j; \quad i, j = 1, 2, \ldots, r;$$

(b) in each experiment one of them certainly occurs—i.e., they contain all elementary events:

$$A_1 + A_2 + \ldots + A_r = I;$$

—then we say that $A_1, A_2, \ldots, A_r$ form a *complete system of events*. For them obviously the following relation holds:

$$P(A_1) + P(A_2) + \ldots + P(A_r) = 1.$$

1.5 CONDITIONAL PROBABILITY

The fundamental notion of *conditional probability* can be illustrated by means of a simple example. Starting from Experiment 3 let us consider the number of yarn breakages ν and the production Z during the same shift. Our aim is to investigate the effect of breakages of the yarn on the quantity of production. For this purpose we raise the question: What is the probability of the event that Z achieves the planned value 1.5 metric tons when ν exceeds a given value, say 20? The probability required is now the probability of the event $A = \{Z > 1.5\}$, but under a certain condition—namely, that the further event $B = \{\nu > 20\}$ takes place. The probability needed is the *conditional probability of event A under the condition that event B also occurs*: this will be denoted by $P(A \mid B)$.

Under the circumstances given in Example 3 the probability of event A has a certain numerical value, and the same is true for the conditional probability of A under condition B. We wish to determine this probability theoretically.

The concept of conditional probability can be reduced to that of ordinary probability. An illustration using the relative frequency of observed values will show how.

Let us follow through the production in n shifts. Assume that the operation of the machine has been found satisfactory in n_A shifts during this period, in other words, the event $A = \{Z > 1.5\}$ (a "good" shift) has occurred n_A times; and that among these n_A shifts there are n_{AB} shifts during which even $B = \{\nu > 20\}$ occurs.

The relative frequency associated with the conditional probability $P(A \mid B)$ is n_{AB}/n_A, which can be written in the form

$$\frac{n_{AB}}{n_A} = \frac{n_{AB}/n}{n_A/n} . \tag{2}$$

But, as we know, the relative frequencies n_A/n and n_{AB}/n fluctuate around the probabilities $P(A)$ and $P(AB)$, respectively. "Substituting" probabilities for the corresponding relative frequencies, we obtain the *definition* of the conditional probability:

$$P(A|B) = \frac{P(AB)}{P(B)} ,$$

which is meaningful, provided that $P(B) > 0$.

Turning to the events of our example, we may write

$$P(Z > 1.5|\nu > 20) = \frac{P(Z > 1.5, \nu > 20)}{P(\nu > 20)} . \tag{3}$$

The numerator gives the probability that the events $\{Z > 1.5\}$ and $\{\nu > 20\}$ occur simultaneously; the corresponding frequency is the number of shifts out of the total number of observed shifts in which both events occur. If the figure obtained for this probability is judged by the management to be too low, then we must alter the conditions of production—i.e., the operation of the machine, the quality of the yarn, etc.—in order to diminish the number of yarn breakages.

1.6 INDEPENDENCE OF EVENTS; THE MULTIPLICATION RULE FOR PROBABILITIES

The concept of independence plays a fundamental role in probability theory and mathematical statistics. Two events are called independent if the probability of one of them is not influenced by whether or not the other occurs; this means for two events A and B that the relation

$$P(A|B) = P(A|\bar{B}) = P(A) \tag{4}$$

holds. From this relation and the definition of conditional probability

$$P(A|B) = \frac{P(AB)}{P(B)}$$

the so-called multiplication rule for independent events follows immediately:

$$P(AB) = P(A)P(B). \tag{4'}$$

As can be easily seen, independence of A and B involves the independence of A and $\bar{B}$, of $\bar{A}$ and B, and of $\bar{A}$ and $\bar{B}$.

Three events may be pairwise independent, but the multiplication rule does not necessarily hold for all three. For complete independence of three events A, B, and C the following relations must hold:

$$\begin{aligned} &P(AB) = P(A)P(B), \qquad &&P(AC) = P(A)P(C), \\ &P(BC) = P(B)P(C), \qquad &&P(ABC) = P(A)P(B)P(C). \end{aligned}$$

Similarly, for the complete independence of r events $A_1, A_2, \ldots, A_r$ the validity of the multiplication rule must hold for each set $A_{i_1}, A_{i_2}, \ldots, A_{i_s}$ $(2 \leqslant s \leqslant r,\ 1 \leqslant i_1 < i_2 < \ldots < i_s \leqslant r)$:

$$P(A_{i_1}A_{i_2}\ldots A_{i_s}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_s}).$$

Since the number of pairs is $\binom{r}{2}$, the number of triplicates is $\binom{r}{3}$ etc., the total number of the relations necessary for the complete independence of r events is

$$\binom{r}{2} + \binom{r}{3} + \ldots + \binom{r}{r} = 2^r - r - 1.$$

We turn now to an example. Let us take a die with its faces labeled from 1 to 6. We denote by η the number on the top side of the die after a throw. η is a random variable; we suppose that it takes each of its possible values with equal probability $\frac{1}{6}$. This means that the probability of the event $A = \{\eta \leqslant 2\}$, for example, can be determined in the following way:

$$P(A) = P(\eta \leqslant 2) = P(\eta = 1) + P(\eta = 2) = 2 \cdot \frac{1}{6} = \frac{1}{3},$$

since different elementary events are always disjoint.

Let us denote by B the event that η is even; the probability of this event is

$$P(B) = P(\eta = 2) + P(\eta = 4) + P(\eta = 6) = 3 \cdot \frac{1}{6} = \frac{1}{2},$$

while the probability of the event that A and B simultaneously occur is

$$P(AB) = P(\eta = 2) = \frac{1}{6} = \frac{1}{3} \cdot \frac{1}{2} = P(A)P(B),$$

which means that the events A and B are independent.

1.7 INDEPENDENCE OF EXPERIMENTS; SAMPLING WITH AND WITHOUT REPLACEMENT

We say that two experiments are independent if the outcome of one experiment does not influence the result of the other. Let us now consider the two experiments as one single experiment, specifying, if necessary, their order of succession and other conditions. If two experiments are independent, then, after their union, each event that is connected with only one of the original experiments is independent of each event that is defined only through the other experiment.

A very important case of the union of independent experiments is that of independent repetitions of the same experiment. This happens in sampling procedures: a sample of size n is the outcome of n independent trials of the same kind. For example, let us assume that Experiment 2 is carried out n times independently and we investigate the diameter X of the cylinders produced. Now if we have specified the random variable X_i as the diameter of the item produced on the ith trial (in order of succession), then the outcome of our new compound experiment will be the random vector variable $(X_1, X_2, \ldots, X_n)$. The corresponding sample space is

$$\mathfrak{X}_n = \{(x_1, x_2, \ldots, x_n) \colon -\infty < x_i < \infty,\ i = 1, 2, \ldots, n\},$$

i.e., the n-dimensional Euclidean space.

We now turn to a comparison of two experiments—one involving the union of two independent experiments, the other, the union of two dependent trials. Let us consider the random choice of items from a finite lot, (a) returning each sample element to the lot before choosing the next sample element, and (b) without such replacement. Let us assume that the lot contains N items, labeled with the integers from 1 through N.

For experiment (a) let us choose randomly two items from the lot, one after the other, returning the first to the lot before choosing the second. Let us denote by η_1 and η_2 the integers attached to the items chosen first and second, respectively.

Now let us determine the probability of choosing a given pair of items on a trial of the experiment; for example, what is the probability of the event $A_2 = \{\eta_1 = 1, \ \eta_2 = 2\}$? Since the choice of the items is performed randomly (e.g., with the aid of random numbers), each of the N possible items has the probability $1/N$ of being chosen for η_1.

In the same way, for the pair (η_1, η_2) each possible pair of integers may occur with probability $1/N^2$. But event A_2 consists of a single elementary event; according to the classical definition of probability given in Section 1.3,

$$P(A_2) = P(\eta_1 = 1, \ \eta_2 = 2) = \frac{1}{N^2}.$$

This means that $P(\eta_1 = 1, \ \eta_2 = 2) = P(\eta_1 = 1)\, P(\eta_2 = 2)$ —i.e., the two choices are independent.

Turning now to experiment (b), the situation is different. In selecting the first item we have N possibilities, but the item chosen on the first trial cannot be chosen on the second. There remain only $N-1$ possibilities for the second selection and altogether $N(N-1)$ for both selections. Since we now have a different experiment from that in (a) we denote the probabilities by P^*, and thus we have

$$P^*(A_2) = P^*(\eta_1 = 1, \eta_2 = 2) = \frac{1}{N(N-1)}.$$

Here the multiplication rule does not hold; thus the two choices are not independent.

The same result is true for each given pair (k, l) $1 \leqslant k, l \leqslant n$; in the case of (a) $k = l$ is possible while, in the case of (b), it is impossible —i.e., $k \neq l$.

Similarly, if we choose n items randomly from a lot of size N, then the probability of a given result $A_n = \{\eta_1 = k_1, \eta_2 = k_2, \ldots, \eta_n = k_n\}$ will be, in the case with replacement,

$$P(A_n) = \frac{1}{N^n},$$

while in the case without replacement it will be

$$P^*(A_n) = \frac{1}{N(N-1)\ldots(N-n+1)}.$$

In determining these probabilities we have had two aims: to point out the difference between independent and dependent trials; and to draw the following consequence of these important relations.

If N is very large as compared with n then—to a good approximation—the following relations hold:

$$\frac{1}{N(N-1)\ldots(N-n+1)} \sim \frac{1}{N^n},$$

that is,

$$P^*(A_n) \sim \frac{1}{N^n} = P(A_n).$$

This means that even in the case without replacement, the choices of single items can be treated as nearly independent of each other, when N is large compared to n. It is very unlikely that an item chosen on one trial will be chosen again in the next trial $\left(\text{in the case of independence the probability of this event is } \frac{1}{N}\right)$.

Finally, we wish to mention that the use of conditional probabilities leads to the same result:

$$P(\eta_1 = 1, \eta_2 = 2) = P(\eta_2 = 2 \mid \eta_1 = 1)\, P(\eta_1 = 1) = \frac{1}{N}\frac{1}{N} = \frac{1}{N^2}$$

the events $\{\eta_1 = 1\}$ and $\{\eta_2 = 2\}$ are independent in the case with replacement, while

$$P^*(\eta_1 = 1, \eta_2 = 2) = P^*(\eta_2 = 2 \mid \eta_1 = 1)\, P^*(\eta_1 = 1) = \frac{1}{N-1}\,\frac{1}{N}$$

is a consequence of the dependence of the two events.

1.8 JOINT EXAMINATION OF TWO MEASUREMENTS OF AN ITEM

As an example of the dependence of two properties let us again consider Experiment 2. We assume that the tolerance limits for the diameter X of the cylinders are 21.88 mm and 22.12 mm respectively, while those for the length Y are 59.85 mm and 60.15 mm. Let us denote the corresponding events by E and F:

$$E = \{21.88 \leqslant X \leqslant 22.12\},$$

$$F = \{59.85 \leqslant Y \leqslant 60.15\}.$$

An item is qualified as a good one if it is found to be good with respect to both properties. This means that a piece is good if the event EF occurs. The probability $P(EF)$ of this event may be written in the form

$$P(EF) = P(F|E)\ P(E) \tag{5}$$

or

$$\begin{aligned} &P(21.88 \leqslant X \leqslant 22.12;\ 59.85 \leqslant Y \leqslant 60.15) \\ &= P(59.85 \leqslant Y \leqslant 60.15 \mid 21.88 \leqslant X \leqslant 22.12)\, P(21.88 \leqslant X \leqslant 22.12) \end{aligned} \tag{5'}$$

We turn now to the corresponding frequencies. Selecting n pieces randomly from the lot produced during a shift, we first classify this sample according to diameter. Let k_E be the number of items whose diameters lie between the given tolerance limits, and k_{EF} be the number of items among these that are also good with regard to length.

Then the relation between the relative frequencies corresponding to Eq. (5) is

$$\frac{k_{EF}}{n} = \frac{k_{EF}}{k_E}\,\frac{k_E}{n}.$$

But, as we know, the probability of the joint event EF can be expressed in the following way, too:

$$\begin{aligned} &P(21.88 \leqslant X \leqslant 22.12;\ 59.85 \leqslant Y \leqslant 60.15) \\ &= P(21.88 \leqslant X \leqslant 22.12 \mid 59.85 \leqslant Y \leqslant 60.15)\, P(59.85 \leqslant Y \leqslant 60.15). \end{aligned}$$

Now classifying the sample items according to the length first, let k_F be the corresponding number of good items. Then the relation expressed by means of the relative frequencies has the form

$$\frac{k_{EF}}{n} = \frac{k_{EF}}{k_F}\,\frac{k_F}{n}.$$

What now if the two events are independent? That is, what if, in the course of production, the length is not influenced by the diameter and vice versa? Expressed in terms of formulae this means that the following relations must be valid:

$$P(21.88 \leqslant X \leqslant 22.12 \mid 59.85 \leqslant Y \leqslant 60.15) = P(21.88 \leqslant X \leqslant 22.12)$$

or

$$\begin{aligned} &P(21.88 \leqslant X \leqslant 22.12;\ 59.85 \leqslant Y \leqslant 60.15)\\ &= P(21.88 \leqslant X \leqslant 22.12)\, P(59.85 \leqslant Y \leqslant 60.15). \end{aligned}$$

The corresponding relations for the relative frequencies are

$$\frac{k_{EF}}{k_F} \approx \frac{k_E}{n}, \quad \frac{k_{EF}}{k_E} \approx \frac{k_F}{n}$$

and

$$\frac{k_{EF}}{n} \approx \frac{k_E}{n}\,\frac{k_F}{n}.$$

Here we have written, instead of equality, approximate equality only since the relative frequencies will show deviations due to random fluctuations. For them—as shown by the first and second relations—independence means that there will be approximately the same proportion of items with good (or bad) lengths among the pieces with good diameters as among those with defective diameters or as among the whole lot. The absolute frequencies of the results of the experiment are summarized in Table 1.

TABLE 1

	Diameter (E)		
Length (F)	Good	Defective	Total
Good	k_{EF}	$k_E - k_{EF}$	k_E
Defective	$k_F - k_{EF}$	$n - (k_E + k_F) + k_{EF}$	$n - k_E$
Total	k_F	$n - k_F$	n

The following two numerical examples illustrate the cases of independence and dependence.

Example 1. The independent case. Suppose we examine $n = 500$ cylinders, and with respect to diameter there are $k_E = 478$ good items among them. Among these 478 items there are 464 pieces that are good with respect to length, too; i.e., they are good items for both properties. Table 2 displays the frequencies of all outcomes in this case.

TABLE 2

	Diameter		
Length	Good	Defective	Total
Good	464	20	484
Defective	14	2	16
Total	478	22	500

The total number of defectives is 36—i.e., the fraction defective is

$$\frac{n - k_{EF}}{n} = \frac{36}{500} = 0.072 = 7.2\%\,.$$

The rate of completely good items is

$$\frac{k_{EF}}{n} = 0.928 = 92.8\%,$$

while

$$\frac{k_{EF}}{k_E} = \frac{464}{484} = 0.967 = 96.7\% \approx \frac{k_F}{n} = \frac{478}{500} = 0.956 = 95.6\%$$

and

$$\frac{k_{EF}}{k_F} = \frac{464}{478} = 0.973 = 97.3\% \approx \frac{k_F}{n} = \frac{484}{500} = 0.968 = 96.8\%\,.$$

It is apparent that in this example the relative frequencies approximately follow the multiplication rule for independent events.

The problem of whether the deviations from the multiplication rule can be regarded as random fluctuations or must be considered as significant (i.e., as being the consequence of dependence) will be treated later under the heading "Testing for Independence".

Example 2. The dependent case. Again let $n = 500$ cylinders be taken randomly from a lot produced during a shift. Suppose that the number of completely good items is again found to be $k_{EF} = 464$, while the number of items that are good with respect to the diameter is $k_E = 470$ and the number of items good in length is $k_F = 468$. Table 3 shows these relations.

The rate of completely good items is again $k_{EF}/n = 464/500 = 0.928 = 2.8\%$. But the rate of good items with respect to diameter among those

TABLE 3

	Diameter		
Length	Good	Defective	Total
Good	464	4	468
Defective	6	26	32
Total	470	30	500

that are good in length is now $k_{EF}/k_F = 464/468 = 0.9915 = 99.15\%$. The rate of good pieces with respect to the length among those that are good in diameter is now $k_{EF}/k_E = 464/470 = 0.9872 = 98.72\%$. The product of these fractions equals $0.9788 = 97.88\%$, which differs essentially from the corresponding value 92.8%.

This means that in this second case the sizes of the diameter and length are not found to be independent: if an item is defective concerning the one size, there is a high probability that the same will hold for the other size too. Obviously, the knowledge about dependence or independence will influence the work done to avoid the bad items in production.

1.9 THE RULE OF TOTAL PROBABILITIES; BAYES' THEOREM IN THE DISCRETE CASE

Let $A_1, A_2, \ldots, A_n$ be a complete system of events in connection with a random experiment—i.e.,

$$A_1 + A_2 + \ldots + A_n = I$$

$$A_iA_j = \emptyset, \; i \neq j, \; i, j = 1, 2, \ldots, n.$$

Let us denote the respective probabilities by

$$P(A_i) \; (i = 1, 2, \ldots, n), \; \sum_{i=1}^{n} P(A_i) = 1.$$

If B is a further event, then the probability of B can be determined in the following way:

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

This relation is the *rule of total probabilities*. The first part is a consequence of the simple fact that

$$BA_1 + BA_2 + \ldots + BA_n = B(A_1 + A_2 + \ldots + A_n) = BI = B$$

and that for $i \neq j$ the events BA_i and BA_j are disjoint, and hence for them the addition theorem holds. The second statement is obtained by applying to each term the definition of conditional probability.

From the basic identity for conditional probabilities

$$\begin{aligned} P(BA_i) &= P(A_i \mid B)\, P(B) \\ &= P(B \mid A_i)\, P(A_i) \end{aligned}$$

it follows that

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{P(B)}.$$

When the rule of total probabilities is substituted in the right-hand side of this relation, the result is known as *Bayes' theorem*:

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)\, P(A_j)}.$$

The system of probabilities $\{P(A_i),\ i = 1, 2, \ldots, n\}$ is called the *prior distribution*, while the system $\{P(A_i \mid B),\ i = 1, 2, \ldots, n\}$, which can easily be seen to add up also to 1 is called the *posterior distribution*.

To reveal the essentials of this theorem we wish to give an example that will show the significance of the "mathematical model" chosen and demonstrate the importance of this theorem in statistical quality control.

Let us turn to Experiment 1, in which a lot of cogwheels consisting of N pieces is considered; the lot contains M defective pieces and $N - M$ good ones. A sample of size n is taken from the lot, and it contains $\varkappa$ defective and $n - \varkappa$ good items. Here $\varkappa$ is a random variable; we assume now that M is unknown but fixed.

Let us suppose that a given lot is delivered from the factory to a motor works, where sampling results in the value $\varkappa = k$.

The management of the works will now put the following question to the statistician: "Knowing the result $\varkappa = k$ of the sampling, what is the probability that the fraction defective M/N in the lot does not exceed a given value p_0 (p_0 small, e.g., 2%)?"

The statistician has to reply: "Having the given information concerning the sample, I am able to examine the question of the fraction defective in the lot [see Chapter III, Section 2.2.1] but I cannot answer your question directly, at least not in the framework of Experiment 1. The event $\{M/N < p_0\}$ has no probability at all because it is not a random event: it is a matter of fact whether this relation is true or not. But let us extend our model in a certain way—i.e., with certain further information or at least under certain assumptions; then I will be in the position to answer the problem."

The extension in question is a model close to Experiment 1. Let us turn therefore to the following.

Experiment 1'. Cogwheels are produced in a factory under stable conditions; from time to time lots of size N are transported to the motor works. There a sample of size n is taken randomly from each lot, and the number $\varkappa_n$ of defective items in the sample is determined.

In this random experiment there are two random quantities: M, the number of defective items in the whole lot; and $\varkappa$, the number of defective items in the sample. Accordingly, we may speak about the probability of the events $A_i = \{M = i\}$ for $i = 0, 1, 2, \ldots, N$, and $B = \{\varkappa = k\}$. In a long run of production and transportation it is possible to determine the relative frequencies belonging to the probabilities $P(A_i)$, $P(B \mid A_i)$ for $i = 0, 1, 2, \ldots, n$. Applying Bayes' theorem, we are able to approximate the probability in question:

$$P(M < N_{p_0} \mid B) = \sum_{i=0}^{Np_0} P(A_i \mid B).$$

This laborious method is not the only possibility; there exists a theoretical approach to this problem, which will be treated in Section 4.2.4.

2 RANDOM VARIABLES

2.1 THE RANDOM VARIABLE AND ITS DISTRIBUTION

2.1.1 The Case of One Variable

As we have pointed out, most quantities occurring in natural science and in engineering are subject to random fluctuations, i.e., they are random quantities. We have attached random experiments to them and have called them random variables. Now we will give a precise definition of the random variable.

First let us survey the random variables according to their possible values. A random variable, as we know, is a quantity that takes on real values randomly. A random vector is an n-tuple of real valued random variables. We would like to remark that the outcome of an observed experiment might be a curve or any other complex geometrical, physical or chemical object or figure that is subject to random variations. These types of random elements will, however, be treated only briefly, in the section on stochastic processes.

A real-valued random variable may take on:

(a) Two possible values. (For example, the random variable X describing the quality of an item might be defined as 1 if the item is defective, 0 if it is good.)

(b) A finite number of possible values. (For example, the number $\varkappa$ of defective items in a lot or in a random sample taken from a lot, the number of items produced in an hour or in a shift, etc.).

(c) Infinitely many discrete values. (In the case of problems arising in practice the number of possible values is finite; nevertheless, frequently they cannot be limited. Besides, it often occurs that an infinite model simplifies the mathematical treatment.)

(d) Any value in an interval of the real axis. (For example, the weight of a chemical substance in a given vessel, the length of the remaining part of a roll of some textile material after cutting off pieces of a given length.)

(e) Any value in a half-infinite interval. (For example, the breaking strength of a ball may vary in the interval $(0, \infty)$; the temperature in a given place of the cosmos may be between absolute zero and infinity; the life span of a part of a machine may take on values in the interval $(0, \infty)$, since it cannot be limited with any value.)

(f) Any value on the real axis $(-\infty, \infty)$. (We shall often use this model for random variables, even if they take on their actual values in a finite interval only. This procedure will be explained later.)

The domains just described, as we know, are the *sample spaces* associated with the random variables.

It is sufficient to consider random variables of the types (*a–f*). Of course, there are many other possibilities, some of which are combinations of the above cases (e.g., two disjoint intervals). The variables in cases (a), (b), and (c) are called discrete, while those in (d), (e), and (f) are called continuous.

In connection with the range of possible values of a random variable it is important to know which values occur frequently and which ones occur only rarely. More precisely: What probabilities are attached to single values or to small intervals? This way of phrasing the question leads to the notion of the probability distribution.

First, let us consider the discrete case. Let X be a random variable associated with a random experiment having the sample space

$$\mathfrak{X} = \{x_1, x_2, \ldots\}.$$

The events $\{X = x_i\}$ are the elementary events, from which all possible further events can be constructed. Consequently, the probability of every compound event can be determined by means of the probabilities of the elementary events:

$$p_i = P\{X = x_i\}, \; i = 1, 2, \ldots . \tag{6}$$

The set of probabilities according to Eq. (6) is called the *probability distribution of the discrete random variable.* (Figure 3.)

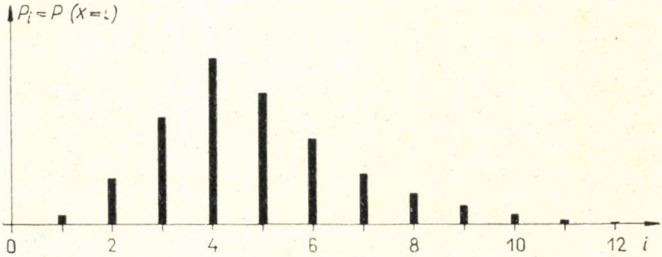


FIG. 3. Discrete distribution

Since the events $\{X = x_i\}$ $i = 1, 2, \ldots,$ form a complete system of events, i.e., their sum is the certain event

$$\{X = x_1\} + \{X = x_2\} + \ldots = I,$$

the corresponding probabilities must add up to 1:

$$P\{X = x_1\} + P\{X = x_2\} + \ldots = p_1 + p_2 + \ldots = 1.$$

We now turn to the continuous case. Let X be a continuous random variable with the sample space

$$\mathfrak{X} = \{-\infty < x < \infty\}.$$

Let us consider events of the form $\{X \leqslant x\}$. For every fixed value of x there exists the probability $P(X \leqslant x)$. This can be approximated by relative frequencies in a long series of trials: we have only to determine the proportion of those trials in which the observed value of X does not exceed x. Since these probabilities depend on the chosen value of x, they may be considered to be a function of x:

$$P(X \leqslant x) = F(x), \quad -\infty < x < \infty.$$

This function $F(x)$ is called the *distribution function of the random variable* X.

Knowing $F(x)$, we are able to determine the probability of further events, e.g., that of the event $\{x_1 < X \leqslant x_2\}$ $(x_1 < x_2)$, namely,

$$\{X \leqslant x_1\} + \{x_1 < X \leqslant x_2\} = \{X \leqslant x_2\};$$

and since the events to be added up are disjoint events, the following relation holds:

$$F(x_1) + P(x_1 < X \leqslant x_2) = F(x_2),$$

from which we obtain

$$P(x_1 < X \leqslant x_2) = F(x_2) - F(x_1).$$

Since the probability of an event is always nonnegative, this relation implies that the distribution function is monotone

$$F(x_1) \leqslant F(x_2) \text{ for } x_1 \leqslant x_2.$$

Further properties of the distribution function will be treated in the next section. In the following we wish to give a mathematical approach to the random variable.

We have started from events of the form $\{X \leqslant x\}$ and obtained by subtraction events according to intervals $\{x_1 < X \leqslant x_2\}$. Addition of disjoint intervals yields events of the form

$$A = \{x_1 < X \leqslant x_2\} + \{x_3 < X \leqslant x_4\} + \dots + \{x_{2k-1} < X \leqslant x_{2k}\},$$

with $x_1 < x_2 \leqslant x_3 < x_4 \leqslant \dots \leqslant x_{2k-1} < x_{2k}$. For the probabilities of these events we have

$$P(A) = [F(x_2) - F(x_1)] + [F(x_4) - F(x_3)] + \dots \\ + [F(x_{2k}) - F(x_{2k-1})].$$

Suppose we continue the procedure of constructing events of new types with the aid of such operations as addition, subtraction, etc. In this way the probabilities of the new events can be obtained if those of the earlier ones are known. Let us denote by $\mathfrak{A}$ the totality of these events. As a matter of fact, each element A of $\mathfrak{A}$ is a part (a subset) of $\mathfrak{X}$, with a well-defined probability $P(A)$.

It should be noted that in the case of a discrete variable, $\mathfrak{A}$ consists of all possible sets of the elementary events $\{X = x_i\}$. It is easy to determine the distribution function of a discrete variable:

$$F(x) = P(X \leqslant x) = \sum_{x_i \leq x} P(X = x_i),$$

i.e., we have to sum probabilities for all x_i not larger than x.

In the case of a continuous random variable it is possible to construct subsets of $\mathfrak{X}$ which are not elements of $\mathfrak{A}$. For example, a subset of the interval (0, 1) can be constructed that has no measure at all; here the measure of a point set agrees with the ordinary length if the set is an interval. We cannot expect that a point set that has no measure will have a probability as an event. For this reason we confine ourselves to certain subsets of $\mathfrak{X}$, to be regarded as the elements and now $\mathfrak{A}$ is defined as the set of those subsets.

Finally, we wish torem ark that the notion of the probability $P(A)$ of an observable event is fundamental; that of the distribution function $F(x)$ is derived from this.

A complete and mathematically satisfactory description of a random variable is as follows: The symbol X stands for a random quantity that may take on values in a sample space $\mathfrak{X}$. The probability that the observed value of X belongs to the event A is $P(A)$. The probability law $P(\,\cdot\,)$ associated with the underlying random experiment is a concept of a more general character in the sense that it can be used to determine the distribution functions, as well as the joint distribution functions, of other random variables belonging to the same experiment.

There is a collection of events $\mathfrak{A}$, every element of which has a certain probability. Events of the special form $\{X \leqslant x\}$ have probabilities $F(x) = P(X \leqslant x)$. On the other hand, the probability of any event can be developed from the distribution function $F(x)$.

Summarizing: A random variable X is defined in terms of a pair $\{\mathfrak{X}, \mathfrak{A}\}$, and with the probability law $P(\,\cdot\,)$ defined on the elements of $\mathfrak{A}$. The distribution function $F(x)$ enables us to determine the probability of each event $A \in \mathfrak{A}$. In most of the cases it will be sufficient to speak about a random variable X with the distribution function $F(x)$.

2.1.2. The Random Variable in Kolmogorov's Theory

The ideas presented in Kolmogorov's original theory of the random variable can be outlined as follows.

Let us consider Experiment 2 (p. 15), i.e., the process of producing cylinders on a certain machine. (This is the random mass phenomenon we shall deal with.) The outcome of the experiment is now not a single (random) quantity but the totality of all quantities by which the piece is completely specified. Suppose the product can be specified by means of a denumerable (infinite) number of quantities $(\xi_1, \xi_2, \ldots, \xi_n, \ldots)$, expressing geometrical, mechanical, thermodynamical, electrical, optical, and other properties of the item. This infinite vector, which we shall denote by ω, can be considered an "elementary event":

$$\omega = (\xi_1, \xi_2, \ldots, \xi_n, \ldots).$$

Upon carrying out an experiment, we come to a result of this type. Note that the elementary event defined earlier is, in the sense of Kolmogorov's theory, a composite one. We will use quotation marks when speaking about "elementary events" in the sense of Kolmogorov's theory, i.e., when we speak of an ω.

Let us denote the set of all possible "elementary events" by Ω, which is called the space of "elementary events".

A subset A of Ω will be treated as an event if we can determine whether A or $\bar{A}$ has occurred in an experiment. The set of all possible events will be denoted by $\mathfrak{A}$. If with respect to a manufactured cylinder the "elementary event" ω $(\omega \in \Omega)$ occurs, then all possible data of the piece are determined. This means that the length X, the diameter Y, and the breaking strength Z, for example, are uniquely determined by ω; in other words they are functions of ω:

$$X = X(\omega),\ Y = Y(\omega),\ Z = Z(\omega).$$

If we now have x for the observed value of the length; $X(\omega) = x$, then this result x does not belong to a single ω, but to a whole set of ω's $\{\omega\colon X(\omega) = x\}$. Thus, this subset of Ω corresponds to an elementary event in $\mathfrak{X}$. Similarly, the set of those "elementary events" for which $X(\omega) < x$, which is a subset of Ω, corresponds to the event $X < x$, which is a subset of $\mathfrak{X}$. Accordingly, in the theory of probability, the starting point is the triplet $\{\Omega, \mathfrak{A}, P\}$. Here Ω is the space of elementary events; $\mathfrak{A}$ is the collection of events A, i.e., of subsets of Ω, for which the probability $P(A)$ is defined. A random variable X is a function defined on

$\Omega: X = X(\omega)$; this must be a "measurable" function, which means the following: the sets ω defined by $\{\omega: X(\omega) < x\}$ are elements of $\mathfrak{A}$, i.e., they have probabilities.

2.1.3. The Case of a Vector Variable

The definition of random variable given in Section 2.1.1 includes the case of vector variables. For the sake of simplicity we shall treat only the two-dimensional case. In this case Z is a vector with two components of real-valued random variable $Z = (X, Y)$.

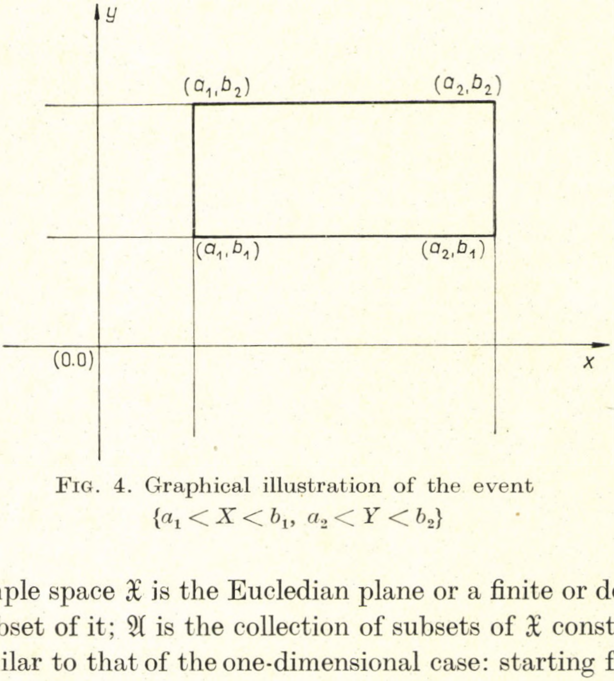


FIG. 4. Graphical illustration of the event $\{a_1 < X < b_1,\ a_2 < Y < b_2\}$

The sample space $\mathfrak{X}$ is the Eucledian plane or a finite or denumerable or other subset of it; $\mathfrak{A}$ is the collection of subsets of $\mathfrak{X}$ constructed in a manner similar to that of the one-dimensional case: starting from events of the form $\{(x, y): -\infty < x < x_0, -\infty < y < y_0\}$ in the plane, we may form by operations (addition, subtraction, etc.) all subsets of the plane, the probability of which can be derived. The distribution function of X is defined now in the form

$$F(x,y) = P(X \leqslant x,\ Y \leqslant y),$$

from which the probabilities of events can be constructed.

As an example we consider the case of the diameter X and the length Y of a cylinder. An item is considered as being good if the requirements

$a_1 < X \leqslant b_1$ and $a_2 < Y \leqslant b_2$ are fulfilled. For the corresponding event

$$A = \{a_1 < X \leqslant b_1,\ a_2 < Y \leqslant b_2\}$$

the following relation holds (see Fig. 4):

$$\{X \leqslant b_1,\ Y \leqslant b_2\} = A + \{X \leqslant b_1,\ Y \leqslant a_2\} + \{X \leqslant a_1,\ a_2 < Y \leqslant b_2\}.$$

Using the definition of the distribution function and recognizing that we have disjoint events,

$$F(b_1, b_2) = P(A) + F(b_1, a_2) + P(X \leqslant a_1,\ a_2 < Y \leqslant b_2).$$

The probability in the last term can be expressed as follows: From the relation for disjoint events,

$$\{X \leqslant a_1,\ Y \leqslant b_2\} = \{X \leqslant a_1,\ a_2 < Y \leqslant b_2\} + \{X \leqslant a_1,\ Y \leqslant a_2\}.$$

It follows that

$$F(a_1, b_2) = P(X \leqslant a_1, a_2 < Y < b_2) + F(a_1, a_2).$$

Finally, we obtain

$$P(a_1 < X \leqslant b_1, a_2 < Y \leqslant b_2) = F(b_1, b_2) - F(a_1, b_2) \\ - F(a_2, b_1) + F(a_1, a_2).$$

The same result is obtained starting from the following relation:

$$\{X \leqslant b_1,\ Y \leqslant b_2\} = A + \{X \leqslant a_1,\ Y \leqslant b_2\} + \{X \leqslant b_1, Y \leqslant a_2\}.$$

Since the events in this relation are not disjoint, the corresponding addition theorem (see Section 1.4) should be applied.

2.2 FURTHER PROPERTIES OF THE DISTRIBUTION FUNCTION THE DENSITY FUNCTION OF A CONTINUOUS RANDOM VARIABLE

The event $\{X < -\infty\}$, being an impossible event, has zero probability; while the event $\{X < \infty\}$—being a certain event—has a probability of one. Furthermore—as we have seen before—the distribution function is a monotone function (Figure 5). Accordingly, any distribution function has the following properties:

(a) $F(x)$ is monotone nondecreasing in $(-\infty, \infty)$;
(b) $F(-\infty) = 0$;
(c) $F(\infty) = 1$.

The distribution function of a continuous random variable is a continuous function. It is constant in an interval (a, b), i.e., $F(a) = F(x) = F(b)$ for $x \in (a, b)$, if the event that the variable takes on some value between a and b has zero probability. For the same reason, a discrete variable has a distribution function that is piecewise constant. It has jumps at those points that are its possible values. Because the value of the distribution function at point x is the probability of the event $\{X \leqslant x\}$, the distribution function is continuous from the right. This means that if $x \to x_0 + 0$, i.e., x tends to x_0 from above, then $F(x) \to F(x_0)$. For example, let X be the random variable that takes on the values 0 and 1 with probabilities $\frac{1}{2}, \frac{1}{2}$: $P(x = 0) = \frac{1}{2}$, $P(x = 1) = \frac{1}{2}$. Then the distribution function is

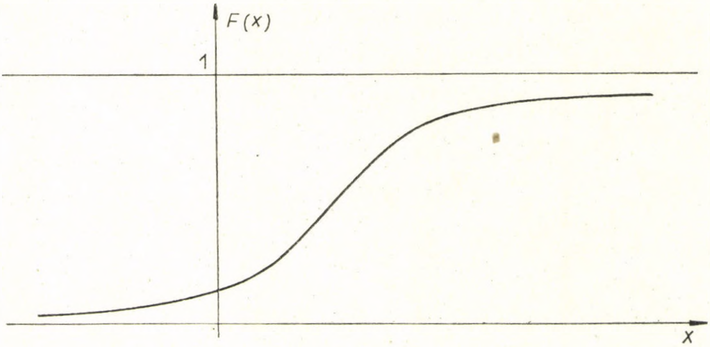


FIG. 5. Continuous distribution function

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{1}{2}, & \text{if } 0 \leqslant x < 1, \\ 1, & \text{if } 1 \leqslant x. \end{cases}$$

This means that in any case $F(x + 0) = F(x)$ and $F(x - 0) \leqslant F(x)$; here the sign $<$ holds if the event $\{X = x\}$ has positive probability.

According to this a distribution function has the property:

(d) $F(x)$ is continuous from the right.

Properties (a)–(d) *characterize* a distribution function.

In the case of a continuous random variable the event $X = x$ has probability 0 for each value of x. For example the event that the diameter X of a cylinder will take on exactly the value 21.9911 . . . mm has zero probability. This is the situation even for all possible single values of X. Therefore the fact that an event has zero probability does not imply that it is an impossible event. Because of this feature of a continuous random variable it is quite irrelevant whether we write $P(X < x)$ or $P(X \leqslant x)$; these values agree.

In the case of a continuous random variable X, the appropriate question is: What is the probability of the event that X will be in a small neighborhood of x, e.g., in the interval $\left(x - \frac{1}{2} \Delta x,\, x + \frac{1}{2} \Delta x\right)$, where Δx is a small positive value? This event has the probability

$$P\left(x - \frac{1}{2} \Delta x < X < x + \frac{1}{2} \Delta x\right) = F\left(x + \frac{1}{2} \Delta x\right) - F\left(x - \frac{1}{2} \Delta x\right).$$

If $F(x)$ has a derivative, then after dividing by Δx and letting Δx tend to zero, we obtain this derivative:

$$\lim_{\Delta x \to 0} \frac{P\left(x - \frac{1}{2} \Delta x < X < x + \frac{1}{2} \Delta x\right)}{\Delta x} = \frac{dF(x)}{dx} = F'(x) = f(x).$$

This means that the following relation holds approximately:

$$P\left(x - \frac{1}{2} \Delta x < X < x + \frac{1}{2} \Delta x\right) \approx f(x)\, \Delta x.$$

The function $f(x)$ is called the *density function of the continuous random* variable X. (Figure 6.) The following relations are obvious:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad \int_{-\infty}^{\infty} f(t)\, dt = 1,$$

$$P(x_1 < X \leqslant x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)\, dt,$$

$$P(X > x) = 1 - F(x) = \int_{x}^{\infty} f(t)\, dt.$$

The density function is nonnegative on the whole real axis: $f(x) \geqslant 0$ $(-\infty < x < \infty)$. The area under its whole curve is 1; the area under the curve belonging to a given interval (a, b) is equal to the probability that X will take on some value of this interval. Furthermore, it is equal to the increase in the distribution function $F(x)$ from a to b.

If the density function is symmetric around a value x_0, then we speak of a symmetric distribution. If $x_0 = 0$, then

$$f(x) = f(-x)$$

for all x, or in the general case

$$f(x_0 - x) = f(x_0 + x)$$

for all values of x. (Figure 7.)

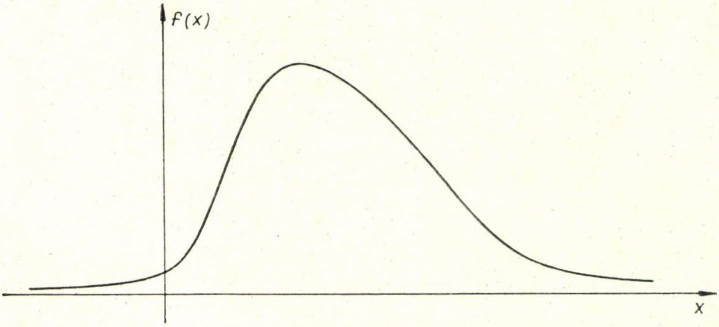


FIG. 6. Density function

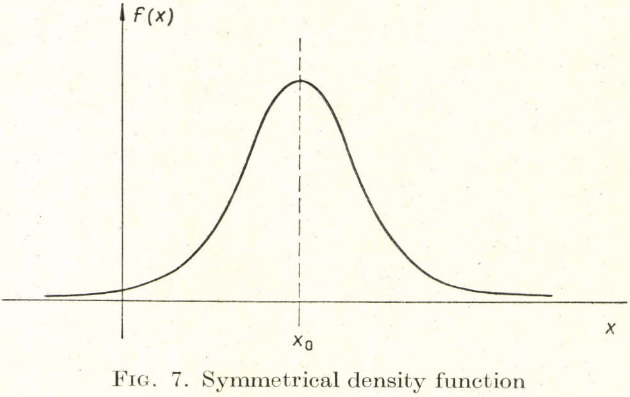


FIG. 7. Symmetrical density function

Let X and Y be two real-valued continuous random variables associated with the same experiment, their joint distribution function being $F(x, y)$. This function has the following simple properties:

(a) $F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geqslant 0$

for any $x_1 \leqslant x_2$, $y_1 \leqslant y_2$,

(b) $F(-\infty, y) = F(x, -\infty) = 0$,

(c) $F(\infty, \infty) = 1$,

as the corresponding events are $\{x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2\}$ in the case (a), the impossible event in the case (b), and the certain event in the case (c). Furthermore

$$F(x, \infty) = P(X \leqslant x) = G(x),$$

$$F(\infty, y) = P(Y \leqslant y) = H(y).$$

$F(x, y)$ is nondecreasing in x for fixed y and in y for fixed x. $G(x)$ is the distribution function of X and $H(y)$ is that of Y; they are called the *marginal distributions.*

The density function is in the two-dimensional case

$$\frac{\partial^2 F(x,y)}{\partial x\, \partial y} = f(x, y).$$

The function $f(x, y)$ is nonnegative and the joint distribution function of X and Y can be expressed by means of it in the following way:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv. \tag{7}$$

Furthermore,

$$P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u, v) du dv.$$

On the basis of the concept of conditional probability introduced in Section 1.5 we may speak of the conditional distribution of X under a certain condition expressed by means of Y. For example, let the condition be $\{Y \leqslant y_0\}$; then

$$P(X \leqslant x \mid Y \leqslant y_0) = \frac{P(X \leqslant x, Y \leqslant y_0)}{P(Y \leqslant y_0)} = \frac{F(x, y_0)}{H(y_0)} = F(x \mid Y \leqslant y_0).$$

$F(x \mid Y \leqslant y_0)$ is a distribution function indeed: if $H(y_0) \neq 0$, then it is nondecreasing; $F(-\infty \mid Y \leqslant y_0) = 0$; $F(\infty \mid Y \leqslant y_0) = 1$.

Now let the condition be $\{y_0 < Y \leqslant y_0 + \varDelta y_0\}$; then

$$F(x \mid y_0 < Y \leqslant y_0 + \varDelta y_0) = P(X < x \mid y_0 < Y \leqslant y_0 + \varDelta y_0)$$
$$= \frac{P(X \leqslant x, y_0 < Y \leqslant y_0 + \varDelta y_0)}{P(y_0 < Y \leqslant y_0 + \varDelta y_0)} = \frac{F(x, y_0 + \varDelta y_0) - F(x, y_0)}{H(y_0 + \varDelta y_0) - H(y_0)}.$$

If Δy_0 tends to 0, we reach finally the condition $\{Y = y_0\}$, which has zero probability. Although the conditional probability has been introduced for conditions with positive probabilities, nevertheless in the present case we may get a conditional probability under a condition with probability zero. Suppose that $H'(y_0) = h(y_0) \neq 0$. Then we may write the above expression in the form

$$F(x \mid y_0 < Y \leqslant y_0 + \Delta y_0) = \frac{\dfrac{[F(x, y_0 + \Delta y_0) - F(x, y_0)]}{\Delta y_0}}{\dfrac{[H(y_0 + \Delta y_0) - H(y_0)]}{\Delta y_0}} .$$

Letting Δy_0 tend to zero and making use of Eq. (7), we obtain the relation

$$F(x \mid Y = y_0) = \frac{\left.\dfrac{\partial F(x, y)}{\partial y}\right|_{y=y_0}}{H'(y_0)} = \frac{\int\limits_{-\infty}^{x} f(u, y_0) du}{\int\limits_{-\infty}^{\infty} f(u, y_0) du} .$$

$F(x \mid Y = y_0)$ is the conditional distribution function of X under the condition $Y = y_0$; this conditional distribution function exists in cases of practical interest. An important application will be given in Section 4.2.4 in Bayes' method in mathematical statistics.

2.3 CHARACTERISTICS OF A RANDOM VARIABLE

To characterize a random variable we have introduced the probability distribution in the case of a discrete random variable and the distribution function for a continuous random variable. These functions are theoretical notions. Knowing the distribution function $F(x)$ of a random variable we also have at our disposal all possible information on the future behavior of the random quantity. For example, in considering Example 2, we can determine to any given high probability—say 0.99—two values a_1 and a_2 such that X will lie between a_1 and a_2 with probability 0.99:

$$P(a_1 < X \leqslant a_2) = 0.99.$$

Therefore we may expect that if the circumstances of the production process do not change, then on the average the diameters of 99% of the items will be found to be between a_1 and a_2.

Though the distribution function contains all information on the fluctuation of the random quantity considered, this information is not in a concise form. In practical life, a better summary of the situation is often needed, some well-chosen values that inform us about important features of the random variable: e.g., a value around which the random variable fluctuates, a measure of the spreading of this fluctuation, etc. In the following we will define some characteristics associated with a random variable. These are expressed with the aid of the distribution function. The connection between these theoretical characteristics and certain empirical quantities calculated from observed values will be shown in the section on statistics.

We introduce now two theoretical quantities characterizing the location of the distribution; these will be followed by two measures of the dispersion of the distribution.

2.3.1 The Expected Value

In the case of a discrete distribution the expected value $E(X)$ of the respective random variable is defined in the following way: If X takes on the values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$, then

$$E(X) = \sum_{(i)} x_i p_i$$

if this sum exists.

The notation $E(\cdot)$ for the expected value will be used throughout.

If X is a continuous random variable with density function $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} t f(t)\, dt$$

if the integral exists.

Let us assume that masses of size p_i are located at the points x_i $(i = 1, 2, \ldots)$ of the real axis, i.e., we have a discrete mass distribution. Then the mass center of this mass distribution is $E(x)$. The same holds in the continuous case if we have a continuous mass distribution on the real line with a mass density $f(x)$ on the point x.

From this analogy, as well as from the mathematical form of the expected value, there follows the simple statement that if a distribution is symmetrical around x_0, then $E(X) = x_0$. If the density function is symmetrical around the origin $x = 0$, then its expected value is 0.

Given a random variable X, we may consider a function $h(x)$ of this variable such as x^2, $\sqrt{|x|}$, e^x, etc. If, for example, X denotes the diameter of a cylinder, then $h(X) = \frac{1}{4} X^2\pi$ is the area of the disk. The quantity $h(X)$ will also show random fluctuations depending on the random fluctuation of X, on the one hand, and on the form of $h(x)$, on the other hand. The expected value of $h(X)$ is

$$E[h(X)] = \int_{-\infty}^{\infty} h(t)\, f(t)\, dt.$$

This means that if the diameter X fluctuates around its expected value $E(X) = \int_{-\infty}^{\infty} tf(t)\, dt$, then the corresponding area will fluctuate around

$$E\left(\frac{1}{4} X^2 \pi\right) = \frac{1}{4} \pi \int_{-\infty}^{\infty} t^2 f(t)\, dt.$$

A simple function of a random variable is its deviation from its expected value: $h(X) = X - E(X)$. This fluctuates around 0:

$$\begin{aligned} E[X - E(X)] &= \int_{-\infty}^{\infty} [t - E(X)]\ f(t)dt = \int_{-\infty}^{\infty} tf(t)dt - E(X) \int_{-\infty}^{\infty} f(t)dt \\ &= E(X) - E(X) = 0. \end{aligned}$$

The fluctuation around the expected value does not imply, however, that the random variable takes on smaller and larger values with equal probability.

2.3.2 Median and Quantiles

The median M of a continuous variable is the value for which the distribution function takes on the value $\frac{1}{2}$:

$$F(M) = \frac{1}{2}$$

or

$$\int_{-\infty}^{M} f(t)dt = \int_{M}^{\infty} f(t)\, dt = \frac{1}{2}.$$

This value is determined uniquely if $F(x)$ is strictly increasing. It may occur, however, that $F(x)$ has the constant value $\frac{1}{2}$ within an interval; in this case any value in this interval may be regarded as the median.

In the case of a discrete random variable the distribution function may jump from a value smaller than $\frac{1}{2}$ to one larger than $\frac{1}{2}$.

Let a be such that $F(a-0) < 0.5 < F(a)$, then a may be considered to be the median.

It is obvious that the symmetry of the distribution implies the coincidence of the median and the expected values.

Considering a continuous random variable only, we introduce the q quantile (sometimes called also $100q$ percentile) $(0 < q < 1)$ as the x value that satisfies the relation

$$F(x_q) = q.$$

For example, $x_{0.5} = M$. The 25 percentile and 75 percentile—called quartiles—are frequently used in practice.

On the average, half of a great number of observed values of the random variable fall between $x_{0.25}$ and $x_{0.75}$.

2.3.3 Variance and Standard Deviation

To characterize the fluctuation of the observed values of X, we introduce the following theoretical quantity, called the variance (dispersion) of X:

$$D^2(X) = E\{[X - E(X)]^2\}.$$

In the case of a continuous random variable the variance takes on the form

$$D^2(X) = \int_{-\infty}^{\infty} [t - E(X)]^2 f(t)dt.$$

In the discrete case, if the possible values of X are $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$, then

$$D^2(X) = \sum_{(i)} [x_i - E(X)]^2 p_i.$$

The standard deviation of the random variable X is the square root of the variance, denoted by $D(X)$.

We shall often use also the notation σ_X^2 or σ^2 for the variance, and σ_X or σ for the standard deviation.

We wish to remark that $D^2(X) = 0$ if and only if X takes on a single value with probability 1, i.e., if X is not subject to random fluctuation at all and is therefore a *constant*.

2.3.4 Mean Deviation

The mean deviation is another measure characterizing the fluctuation of the observed values of X; it is defined as the expected value m of the absolute deviation $|X - E(X)|$:

$$m = E[|X - E(X)|].$$

In the discrete case this has the form

$$m = \sum_{(i)} |x_i - E(x)| p_i,$$

while in the continuous case,

$$m = \int_{\infty}^{\infty} |t - E(X)| \quad f(t)\,dt.$$

The random variable $|X - E(X)|$ fluctuates around the mean deviation.

The variance as a measure of dispersion has the great advantage that it is easy to manage mathematically. Nevertheless, the mean deviation is often applied, particularly in the textile industry.

2.3.5 Moments

The characteristics given above are special cases of moments.

The rth *moment* of a random variable is defined as the expected value of X^r, which in the discrete case has the form

$$\mu_r = \sum_{(i)} x_i^r p_i,$$

and in the continuous case,

$$\mu_r = \int_{-\infty}^{\infty} t^r \; f(t)dt.$$

In the case of distributions symmetric around the origine, for odd values of r the moments vanish: $\mu_{2k+1} = 0$, $k = 1, 2, \ldots$

The rth *absolute moment* is the expected value of $|X|^r$:

$$m_r = \sum_{(i)} |x_i|^r p_i \quad \text{and} \quad m_r = \int_{-\infty}^{\infty} |x|^r \quad f(t)dt$$

in the discrete and continuous cases, respectively.

The *central moments* are defined in a similar way, forming the expected values of $[X - E(X)]$ instead of X, i.e., they are the absolute moments of the random variable $X' = X - E(X)$. For example, the rth absolute central moment is, in the continuous case,

$$C_r = \int_{-\infty}^{\infty} |t - E(X)|^r f(t)\, dt\,.$$

In particular, $C_2 = D^2(X)$.

Further characteristics of the distribution can be derived from the different moments. The measure of the asymmetry of a density curve is called the skewness; denoting it by γ_1, it is defined as

$$\gamma_1 = \frac{E\{[(X-E(X)]^3\}}{D^3(X)}.$$

The *curtosis* γ_2 is a measure of the *flatness* of the density curve; the definition of this measure is

$$\gamma_2 = \frac{E\{[X-E(X)]^4\}}{D^4(x)}\,.$$

The reader is cautioned when using these measures; cf. [A 28], pp. 85–86, 92–93.

2.4 INDEPENDENCE OF RANDOM VARIABLES

2.4.1 The Notion of Independence

In Section 1.6 we introduced the concept of the independence of events. Before giving a precise definition of the independence of random variables, we turn to an example.

Let the experiment be the production of cylinders on two machines, denoted by M_1 and M_2, respectively. Consider the pair of random variables (X, Y), where X is a diameter produced by M_1 and Y is the diameter simultaneously produced by M_2. In general, the two diameters are indepedent: the observed value of X_1 will not be influenced by that of Y_1, the diameter of the other item produced on M_2 at the same time. Dependence might arise in this process in the following way: Both machines are fed by the same energy source, and irregular fluctuations in the electric current may influence the formation of the items.

As a second example consider the diameter X and the hardness Y of a steel cylinder. These variables may be independent or dependent according to the method of production; we may assume independence if the procedure consists of cutting, but dependence if pressing equipment is used.

The independence of random variables is based on the independence of events. The independence of X and Y explained above in the first example means that for all values $x_1, x_2(> x_1), y_1, y_2(> y_1)$ the events $\{x_1 < X \leqslant x_2\}$ and $\{y_1 < Y \leqslant y_2\}$ are independent. Therefore, in the case of independence, according to the multiplication rule, the relation

$$P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2) = P(x_1 < X \leqslant x_2) \quad P(y_1 < Y \leqslant y_2)$$

must hold.

It can be easily seen that for the independence of the random variables X and Y it is sufficient to satisfy the relation $P(X \leqslant x, Y \leqslant y) = P(X \leqslant x) P(Y \leqslant y)$ for all x and y.

Denoting the joint distribution by $F(x, y)$ and the marginal distribution functions of X and Y by $G(x)$ and $H(y)$, respectively, this means that

$$F(x, y) = G(x)H(y) \text{ for all } x \text{ and } y.$$

The same property can be proved for the corresponding joint density:

$$f(x, y) = g(x)\, h(y). \tag{8}$$

The independence of more than two random variables is analogously defined. For example, the three random variables X_1, X_2, X_3 are independent, if for all real values x_1, x_2, x_3 the following relation holds:

$$P(X_1 \leqslant x_1, X_2 \leqslant x_2, X_3 \leqslant x_3) = P(X_1 \leqslant x_1) \; P(X_2 \leqslant x_2) \; P(X_3 \leqslant x_3).$$

The concept of independence plays a basic role in statistical sampling. As we will see, a statistical sample is a given number of observations on a random variable X. These observations $X_1, X_2, \ldots, X_n$ will be considered as n different random variables, and they will be assumed in general to be completely independent. In most cases a sample consisting of independent elements is more informative than a sample containing dependent elements. In reality, however, the sample elements are often not perfectly independent: e.g., the items produced one after the other on the same machine may show some slight dependence. In any case, if it is possible, the sampling procedure is to be carried out in a way which ensures that the sample elements can be considered to be independent.

2.4.2 Correlation Coefficient

Let X and Y be random variables with the joint distribution function

$$P(X \leqslant x, Y \leqslant y) = F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv.$$

For a pair of random variables, or in general for more variables, it is customary to introduce the so-called mixed moments. It suffices in our treatment to regard the first central mixed moment, i.e., the expected value of the product of $X - E(X)$ and $Y - E(Y)$:

$$m_{XY} = E\{[X - E(X)]\ [Y - E(Y)]\} =$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [u - E(X)]\ [v - E(Y)]\ f(u, v)\ du dv.$$

m_{XY} is called the *covariance* of X and Y. In the independent case this quantity is zero; this is a simple consequence of Eq. (8), i.e., the factorization of the joint density function, applied to the variables $[X - E(X)]$ and $[Y - E(Y)]$. It may happen, however, that $m_{XY} = 0$ holds although X and Y are not independent. Therefore, if we know only that $m_{XY} = 0$, we say that *the random variable X and Y are uncorrelated.*

The correlation coefficient $\varrho(X, Y)$ of X and Y is defined as follows:

$$\varrho(X, Y) = \frac{E\{[X - E(X)]\,[Y - E(Y)]\}}{D(X)\ D(Y)} = \frac{m_{XY}}{\sigma_X \sigma_Y},$$

where σ_X and σ_Y denote the standard deviations of X and Y respectively. According to the well-known inequality of Schwarz, the value of $\varrho(X, Y)$ lies between -1 and $+1$:

$$-1 \leqslant \varrho(X, Y) \leqslant +1.$$

In a certain sense $\varrho(X, Y)$ measures the strength of the dependence between the random variables X and Y.

If $m_{XY} = 0$, then $\varrho(X, Y) = 0$ as well; that is, X and Y are uncorrelated, which occurs always if they are independent. *If, however, X and Y are jointly normally distributed*, i.e., if they have a joint density function of the form

$$f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y (1 - \varrho^2)^{1/2}} \times$$
$$\exp\left[-\frac{1}{2(1 - \varrho^2)}\left(\frac{(x - a)^2}{\sigma_X^2} - 2\varrho\frac{(x - a)\,(y - b)}{\sigma_X \sigma_Y} + \frac{(y - b)^2}{\sigma_Y^2}\right)\right],$$

then $\varrho(X, Y) = 0$ implies the independence of X and Y. This follows from the fact that the parameter ϱ in the density function agrees with $\varrho(X, Y)$;

if it is zero, then the density function will be the product of the two Gaussian density functions belonging to X and Y.

In the case of $\varrho(X, Y) = 1$ or -1, *between the random variables X and Y* a strict linear connection exists:

$$Y = aX + b,$$

where $a > 0$ if $\varrho(X, Y) = +1$, while $a < 0$ if $\varrho(X, Y) = -1$.

Even if the correlation coefficient cannot give an exact proof of the independence of the variables (except when they are jointly normally distributed), it is a very useful tool in determining the existence of dependence and as a measure of the dependence.

2.5 DISTRIBUTION OF FUNCTIONS OF RANDOM VARIABLES

Having a random variable X, it often happens that we want to investigate a function $h(X)$ of X. For example, given the diameter X of a ball, we may be interested in the volume $h(X) = \frac{4}{3}(X/2)^3\pi = \frac{1}{6}X^3\pi$.

Given two or more random variables, we are often interested only in their sum or product. For this reason it is of great importance to know how to determine the distribution of functions of random variables.

2.5.1 Functions of a Single Random Variable

The determination of the distribution of $h(X)$ is simple if $h(X)$ is a strictly monotone function. We shall show this in a few examples. First we will restrict ourselves to continuous random variables.

Let $P(X \leqslant x) = F(x)$ and $Y = aX + b$ with $a > 0$. Then the following steps are obvious:

$$G(x) = P(Y \leqslant x) = P(aX + b \leqslant x) = P\left(X \leqslant \frac{x-b}{a}\right).$$

But according to the definition, we have for the distribution function

$$G(x) = F\left(\frac{x-b}{a}\right),$$

and for the density function

$$G'(x) = g(x) = \frac{1}{a}\; f\left(\frac{x-b}{a}\right).$$

If $a < 0$, then—after dividing by a—the inequality sign will change and we obtain

$$G(x) = P\left(X > \frac{x-b}{a}\right) = 1 - F\left(\frac{x-b}{a}\right),$$

and

$$g(x) = \frac{1}{|a|} f\left(\frac{x-b}{a}\right).$$

Let the random variable X be positive: then for $h(x) = \sqrt{x}$ we have

$$G(x) = P(\sqrt{X} < x) = P(X < x^2) = F(x^2),\ g(x) = 2xf(x^2).$$

If $h(x)$ is not monotone some precaution must be taken. Again let $P(X \leqslant x) = F(x)$ and $h(x) = x^2$. Then for $x \geqslant 0$

$$G(x) = P(X^2 \leqslant x) = P(-\sqrt{x} \leqslant X \leqslant \sqrt{x}) = F(\sqrt{x}) - F(-\sqrt{x}),$$

and for the density function

$$g(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} [f(\sqrt{x}) - f(-\sqrt{x})] & \text{if} \quad x \geqslant 0, \\ 0 & \text{if} \quad x < 0. \end{cases}$$

Finally, let X be a discrete random variable with possible values $x_1, x_2, \ldots$ and with corresponding probabilities $p_1, p_2, \ldots$. Then the possible values of $h(x)$ are $y_1 = h(x_1)$, $y_2 = h(x_2), \ldots$. The distribution function of $h(x)$ is given by

$$P[h(X) < y] = \sum_{h(x_i)<y} P(X = x_i) = \sum_{h(x_i)<y} p_i$$

and the probability distribution of $h(X)$ is

$$P[h(X) = y] = \sum_{h(x_i)=y} P(X = x_i) = \sum_{h(x_i)=y} p_i.$$

2.5.2 Sum of Independent Random Variables

Let X and Y be continuous random variables with distribution functions

$$F(x) = \int_{-\infty}^{x} f(t)dt \text{ and } G(x) = \int_{-\infty}^{x} g(t)dt.$$

In order to determine the distribution function of the random variable $Z = X + Y$, we consider first the case when Y is fixed: $Y = v$. Under this condition the event $\{Z = X + Y \leqslant z\}$ agrees with the event $\{X \leqslant z - v\}$. The probability of this event equals $F(z - v)$. On the other hand, the probability of the event that Y falls in a small neighborhood of the value v is $g(v)\,\Delta v$. Now the two events, namely $\{X \leqslant z - v\}$ and $\left\{v - \frac{1}{2}\Delta v < Y \leqslant v + \frac{1}{2}\Delta v\right\}$, are independent; consequently, the probability of their simultaneous occurrence is the product of the, probabilities

$$F(z - v)g(v)\Delta v. \tag{9}$$

Let us now divide the real axis $(-\infty < v < \infty)$ into disjoint intervals of length Δv.

The events of Y taking on values from the different intervals are disjoint. Therefore, the probability of the event $\{X + Y < z\}$ is obtained by adding up the probabilities given by Eq. (9) for all intervals of this division and by letting Δz tend to zero; because of the notion of the definite integral this sum tends to the expression

$$\int_{-\infty}^{\infty} F(z - v)g(v)dv.$$

Similarly we may obtain for the same probability the equal quantity

$$\int_{-\infty}^{\infty} G(z - v)f(v)dv.$$

The density function can be obtained by taking the derivative:

$$\int_{-\infty}^{\infty} f(z - v)\;\; g(v)dv = \int_{-\infty}^{\infty} g(z - v)f(v)dv.$$

Turning now to the case of discrete variables, let X and Y be independent random variables with possible values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, respectively. Let us denote the corresponding probabilities by $P(X = x_i) = p_i$ and $P(Y = y_j) = q_j$ for $i, j = 1, 2, \ldots$. As can be seen, the following relation holds for the probability distribution of the sum $Z = X + Y$:

$$P(Z = z) = \sum_{x_i + y_j = z} p_i q_j\,,$$

i.e., summation is to be carried out for those pairs (i, j) for which $x_i + y_j =$

$= z$. If, for example, $x_i = y_i = i-1$, i.e., the possible values of both variables are the non-negative integers, then

$$P(Z = k) = \sum_{i=0}^{k} p_i q_{k-i}, \; k = 0, 1, 2, \ldots$$

Returning to the continuous case, the same result may be obtained in the following way: The event $\{X + Y < z\}$ consists of the set of points in the (x, y) plane for which $x + y < z$. To obtain the probability of the event $\{X + Y < z\}$ we have to integrate the joint density function $h(x, y)$ of the random variables X and Y on this domain:

$$P(X + Y < z) = \int\limits_{u+v<z}\int h(u, v)\,du\,dv = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-u} h(u, v)\,dv \right) du.$$

This formula is valid for dependent random variables as well.

If the random variables X and Y are independent with density functions $f(u)$ and $g(v)$ respectively, then

$$h(u, v) = f(u)\,g(v)$$

and

$$P(X + Y < z) = \int_{-\infty}^{\infty} f(u) \left[\int_{-\infty}^{z-u} g(v)\,dv \right] du = \int_{-\infty}^{\infty} f(u)\,G(z - u)\,du,$$

where $G(y)$ is the distribution function of Y, which is the result obtained previously.

2.5.3 Product and Quotient of Two Independent Random Variables

Let X and Y be independent random variables with density function $f(x)$ and $g(y)$, respectively; $h(x, y) = f(x)\,g(y)$.

Applying the consideration used at the end of the previous section we obtain, for the distribution of the product,

$$P(XY \leqslant z) = \int\limits_{uv<z}\int h(u, v)\,du\,dv\,,$$

and, for the distribution of the quotient,

$$P\left(\frac{X}{Y} \leqslant z\right) = \int\limits_{u<vz}\int h(u, v)\,du\,dv\,.$$

Taking into consideration the fact that X and Y may be both positive and negative, we obtain

$$\int_{-\infty}^{\infty} g(v) f\left(\frac{z}{v}\right) dv$$

for the density function in the case of the product, and

$$\int_{-\infty}^{\infty} |v| g(v)\ f(v)\ dv$$

in the case of the quotient.

2.6 SOME PROPERTIES OF THE EXPECTED VALUE AND OF THE VARIANCE

Here wo consider some simple properties of the two most important characteristics of a random variable, without giving any proof. We have already met some of these properties implicitly before. The relations given below are valid for both discrete and continuous random variables.

Let X and Y be random variables associated with an experiment and let a be a real-valued constant. Then for the expected value the following relations hold:

$$\text{(I)} \qquad E(X + a) = E(X) + a,$$

$$\text{(II)} \qquad E(aX) = a\,E(x),$$

$$\text{(III)} \qquad E(X + Y) = E(X) + E(Y),$$

the last also being valid for more than two random variables.

If X and Y are *independent random variables*, then

$$\text{(IV)} \qquad E\,(XY) = E(X)\,E(Y).$$

No similar rule exists for further operations in general. For example — in general — $E(X/Y)$ does not equal $E(X)/E(Y)$, even if X and Y are independent.

Let us turn now to the variance and standard deviation for which the following relations hold:

$$\text{(V)} \qquad D(X + a) = D(X),$$

$$\text{(VI)} \qquad D(aX) = |a|\ D(X),$$

$$\text{(VII)} \qquad D^2(X + Y) = D^2(X) + D^2(Y) + 2D(X)\,D(Y)\,\varrho(X, Y).$$

If X and Y are *independent* or *uncorrelated*, i.e., $\varrho(X, Y) = 0$, then

$$D^2(X + Y) = D^2(X) + D^2(Y). \tag{VII'}$$

This relation yields for the standard deviation

$$D(X + Y) = [D^2(X) + D^2(Y)]^{1/2}.$$

The practical importance of the last relation will be clear from the following example: Cylinders of two different lengths are produced; let the nominal values be 12 and 28 mm, respectively. The cylinders have to be fitted in order to obtain a cylinder of length 40 mm. According to previous experience, the smaller cylinder is produced with standard deviation 0.25 mm, the larger with 0.12 mm. If the two production processes are independent and the fitting is carried out by random selection, then the standard deviation of the joint cylinder will be, not $0.05 + 0.12 = 0.17$ mm, but only $(0.05^2 + 0.12^2)^{1/2} = 0.13$ mm.

A further important consequence of Eq.(VII) is the following: Let $X_1, X_2, \ldots, X_n$ be identically distributed independent random variables, with the common variance $D^2(X_i) = D^2$ $(i = 1, 2, \ldots, n)$. Denoting by $\bar{X}$ the arithmetic mean of the random variables, i.e.,

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n},$$

the variance of $\bar{X}$ is

$$D^2(\bar{X}) = D^2\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{D^2}{n} \tag{VIII}$$

and hence its standard deviation is

$$D(\bar{X}) = \frac{D}{\sqrt{n}}. \tag{VIII'}$$

This means that the standard deviation of the arithmetic mean of a sample is inversely proportional to the square root of the sample size.

2.7 THEOREM OF CHEBYSHEV AND RELATED INEQUALITIES

Sometimes we have no more information on a random variable than its expected value and standard deviation. Even in this case we may say something about its random fluctuation in the following sense. Let X be a random variable, discrete or continuous, having the expected value

$E(X)$ and standard deviation $D(X)$. For the deviation of X from its expected value the so-called Chebyshev inequality holds:

$$P[|X - E(X)| > \beta D(X)] \leqslant \frac{1}{\beta^2}$$

or equivalently

$$P[|X - E(X)| \leqslant \beta D(X)] \geqslant 1 - \frac{1}{\beta^2}.$$

This relation is true for all $\beta > 0$, but it is reasonable only for $\beta > 1$. The advantage of this inequality is its validity for all possible distributions for which the expected value and variance exist. From this theorem we can conclude, for example, that for any random variable the probability that it will deviate from its expected value by more than three times its standard deviation is smaller than 1/9.

A stronger inequality has been obtained by Gauss for *continuous* random variables; this we formulate when the density function has a unique maximum (the so-called mode) that coincides with the expected value: $f(x) \leqslant f[E(X)]$ for all x. This is satisfied for symmetric random variables having a unique mode. For these the following inequality is valid:

$$P[|X - E(X)| > \beta D(X)] \leqslant \frac{4}{9\beta^2}.$$

This relation is sometimes referred to as the Camp–Meidell inequality.

TABLE 4

β	$\frac{1}{\beta^2}$	$\frac{4}{9\beta^2}$	Normal distribution
2.0	0.25	0.1111	0.0455
2.5	0.16	0.0711	0.0124
3.0	0.111	0.0493	0.0027
3.5	0.082	0.0364	0.0005
4.0	0.062	0.0275	0.0001

For some values of β Table 4 gives the upper limit of the probability that X falls outside the interval

$$[E(X) - \beta D(X),\ E(X) + \beta D(X)]$$

in the respective cases. The last column contains the exact probabilities for the same event if X is distributed according to the so-called Gaussian law, i.e., distributed normally. This distribution will be introduced in the next section.

2.8 LAWS OF LARGE NUMBERS

From Chebyshev's inequality, the laws of large numbers can be derived. The classical law of J. Bernoulli justifies the introduction of the notion of probability as "the value around which the random fluctuation of the relative frequencies occurs".

Let A be an event associated with an experiment with probability $p = P(A)$. Let us denote by $\varkappa_n$ the frequency of the event A in a sequence of trials of size n: for any small $\varepsilon > 0$ the probability of the event $\{|(\varkappa_n/n) - p| > \varepsilon\}$ tends to 0 as n increases:

$$\lim_{n\to\infty} P\left(\left|\frac{\varkappa_n}{n} - p\right| > \varepsilon\right) = 0.$$

We express this fact in the following way: The relative frequency of an event in a sequence of trials converges *stochastically* to the probability of the event as the number of trials tends to infinity. This ensures that for large enough n the relative frequency very rarely deviates from the probability by more than a given small value ($\varepsilon > 0$).

A similar statement holds for the arithmetic mean of random variables. Let the random variables $X_1, X_2, \ldots$ be pairwise independent and identically distributed with expected value $E = E(X_i)$, $i = 1, 2, \ldots$; then the random variable

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

tends stochastically to the expected value E as n increases:

$$P(|\overline{X}_n - E| > \varepsilon) \to 0 \text{ if } n \to \infty.$$

These theorems ensure the possibility of a good approximation or, as the terminology of mathematical statistics calls it, *estimation* of theoretical quantities, such as, for example, the probability of an event, the expected value of a random variable, etc.

Finally, we wish to remark that for the relative frequency and for the arithmetic mean of random variables an even stronger statement holds: *the strong law of large numbers*. According to this theorem, the arithmetic mean and the relative frequency tend with probability 1 to the expected value and the probability of the event, respectively. We shall not consider the precise meaning of this statement.

2.9 SOME IMPORTANT TYPES OF PROBABILITY DISTRIBUTION

Random variables occurring in industry and the natural sciences are often subject to random fluctuations that show certain regularities, i.e., the distribution of the random variable has a well-defined form. In a given practical case this distribution is of a certain type—or family—of distributions. It has a given analytical form in which there is a finite number of parameters. Let, for example, X be the life span of a bulb from a certain production process; long experience has shown that its random fluctuations are subject to the following distribution law:

$$P(X \leqslant x) = 1 - e^{-ax^b}, \quad a > 0,\ b \geqslant 1.$$

Given a lot of bulbs produced under fixed conditions, the parameters a and b will have well-defined constant values. For a lot produced during the next shift the type of distribution will be the same, though a and b may have different values.

Accordingly, in certain situations we will have to deal with parametric sets of distributions; in the above case this set is the following:

$$\{F(x; a, b) = 1 - e^{-ax^b}, \quad 0 < a < \infty,\ 1 \leqslant b < \infty\}.$$

2.9.1 Gaussian or Normal Distribution

The fluctuations of the random variable X are subject to the *normal* or *Gaussian law* with parameters a and σ $(-\infty < a < \infty,\ \sigma > 0)$ if it has the following distribution function:

$$P(X \leqslant x) = \Phi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} \exp\left(-\frac{(t-a)^2}{2\sigma^2}\right) dt, \quad -\infty < x < \infty.$$

Hence the distribution function is uniquely determined by a given pair (a, σ). The set of possible values for the parameters $\{(a, b)\colon -\infty < a < \infty,\ \sigma > 0\}$ is called the *parameter space*. (Figure 8.) For the density function we have

$$\frac{d\Phi(x; a, \sigma)}{dx} = \varphi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

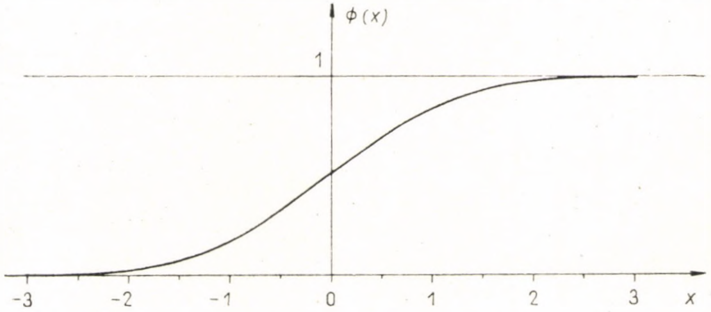


FIG. 8. Distribution function of the standard normal variable

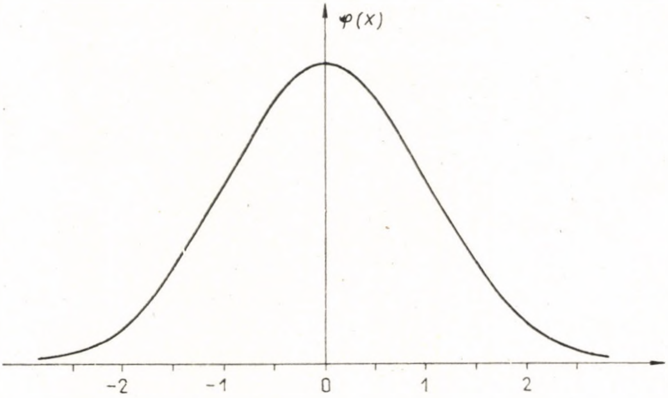


FIG. 9. Density function of the standard normal variable

It can be seen that the parameters a and σ agree with the expected value and the standard deviation of X, respectively:

$$E(X) = \int_{-\infty}^{\infty} t\,\varphi(t; a, \sigma)\,dt = a,$$

$$D^2(X) = \int_{-\infty}^{\infty} (t - a)^2\,\varphi(t; a, \sigma)\,dt = \sigma^2 .$$

We shall denote the distribution and density function of a normally distributed random variable having parameters a (expected value) and σ (standard deviation) throughout the text by $\Phi(x; a, \sigma)$ and $\varphi(x; a, \sigma)$, and the distribution itself by $N(a, \sigma)$. (Figure 9.) The random variable with the distribution $N(0, 1)$ will be called the *standard normal variable.* If X has the distribution $N(a, \sigma)$ then the random variable

$$X' = \frac{X-a}{\sigma}$$

has distribution $N(0, 1)$; X' is called the standardized form of the variable X.

The distribution and density function of the standard normal variable will be denoted by $\Phi(x)$ and $\varphi(x)$, respectively. They are of the form

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt, \qquad -\infty < x < \infty,$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \qquad -\infty < x < \infty.$$

The values of these functions are tabulated in Tables IV (p. 385) and III (p. 384). It is enough to give tables for $x > 0$, since $\varphi(x)$ is symmetrical around the origin and hence $\varphi(-x) = \varphi(x)$; it can easily be seen that $\Phi(-x) = 1 - \Phi(x)$.

Let X be distributed according to the law $N(a, \sigma)$ (a and σ known). In order to be able to determine the probability of the event $\{b_1 \leqslant x \leqslant b_2\}$ with the aid of Table IV, we have to write it in the form

$$\left\{\frac{b_1-a}{\sigma} \leqslant \frac{x-a}{\sigma} \leqslant \frac{b_2-a}{\sigma}\right\};$$

the random variable now standing in the middle follows the standard normal law and therefore

$$P(b_1 \leqslant x \leqslant b_2) = \Phi\left(\frac{b_2-a}{\sigma}\right) - \Phi\left(\frac{b_1-a}{\sigma}\right).$$

We shall often need the probabilities of intervals symmetric to the expected value, measured in units of the standard deviation; i.e., the probabilities of events of the form $\{a - \lambda\sigma < x < a + \lambda\sigma\}$ for certain $\lambda > 0$. For these events we have

$$P(a-\lambda\sigma < X < a+\lambda\sigma) =$$
$$P\left(-\lambda < \frac{X-a}{\sigma} < \lambda\right) = \Phi(\lambda) - \Phi(-\lambda) = 2\,\Phi(\lambda) - 1.$$

For some often-used values of λ the values of these probabilities are the following:

λ	$2\Phi(\lambda) - 1$
1	0.68269
2	0.95450
3	0.99730
4	0.99994

These figures show, for example, that the interval $(a - 3\sigma, a + 3\sigma)$ of total length 6σ contains 99.73% of the distribution. This means that out of 1000 observations made on X, 997 will be within this interval and only two to three outside it on the average.

A further remark of great importance is the following: The probability of the event $\{X < a - 4\sigma\}$ is 0.00003; i.e., out of 100,000 observations on X we will find values smaller than $a - 4\sigma$ only three times on the average. This means that though the normal variable X may take on all positive and negative values, it takes on values smaller than a certain value with very small, even negligible, probability. In the same way there is only a negligibly small probability that X will be outside a certain interval depending on a and σ, e.g., $(a - 4\sigma, a + 4\sigma)$. Consequently, if we have in practice a random variable that is positive or that ranges over a finite interval only, but whose random fluctuations show a close relation to a normal variable, then in many cases it may be considered a normal variable. This circumstance is of great theoretical and practical advantage in solving various problems.

The statement that the random fluctuation of X is close to the normal law will be treated more rigorously in the section on testing statistical hypotheses. As a first rough examination of this question, however, we will consider n observations on X. We count the number of observed values falling in the intervals $(a - \sigma, a + \sigma)$, $(a - 2\sigma, a + 2\sigma)$, $(a - 3\sigma, a + 3\sigma)$ and then compare them with the values $0.67n$, $0.95n$, $0.997n$, respectively. Since a and σ are unknown, we may replace them by values computed from the observed data.

According to central limit theorems, under mild conditions the distribution of the sum of independent random variables tends to the normal law rather rapidly. Therefore it is justifiable to apply the Gaussian distribution in many practical cases where the sum of random variables that are not necessarily normally distributed occurs.

The results of measurements carried out with different types of measuring instruments behave as normal variables to a good approximation. Similarly, Gaussian distributions occur in many cases of serial production. But *not in all cases:* the distribution must be statistically examined. If, for example, the diameter X of a steel ball is subject to the law of normal distribution, then its volume and weight, which are proportional to X^3, cannot have a normal distribution.

Let X_i, $i = 1, 2, \ldots, n$ be independent random variables distributed according to $N(a_i, \sigma)$, $i = 1, 2, \ldots, n$. Then the sum $Y = \sum_{i=1}^{n} X_i$ also has a Gaussian distribution, according to relations (III) and (VII′) in Section 2.6, with mean value $\sum_{i=1}^{n} a_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$. In particular, if the independent random variables $X_1, X_2, \ldots, X_n$ have the common distribution $N(a, \sigma)$ then $Y = \sum_{i=1}^{n} X_i$ is distributed according to $N(na, \sqrt{n}\,\sigma)$ and $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ has the distribution $N(a, \sigma/\sqrt{n})$.

2.9.2 The Poisson Distribution

The discrete random variable X with possible values $0, 1, 2, \ldots,$ is distributed according to the Poisson law if its probability distribution is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots,$$

for some value of the constant parameter $\lambda > 0$.

λ is simultaneously the expected value and the variance of the random variable:

$$E(X) = \lambda, \qquad D(X) = \sqrt{\lambda}\,.$$

Using the formula given for the distribution of the sum of independent random variables in Section 2.5, we obtain the following result (Fig. 10).

Let $X_1, X_2, \ldots, X_s$ be independent random variables having Poisson distribution with parameters $\lambda_1, \lambda_2, \ldots, \lambda_s$ respectively. Then their sum $\sum_{i=1}^{S} X_i$ is also distributed according to the Poisson law with parameter $\sum_{i=1}^{S} \lambda_i$.

The fluctuations of many random quantities occurring in nature and in industrial practice follow the Poisson distribution; e.g., the number of disintegrated particles in a radioactive substance disintegrating during a given time interval, the number of colloid particles in a liquid of a given quantity, the number of air bubbles in a continuous homogeneous material (such as rough-cast glass), the number of yarn breakages occurring on a machine during a shift, the number of flaws on a given length of a textile roll, etc.

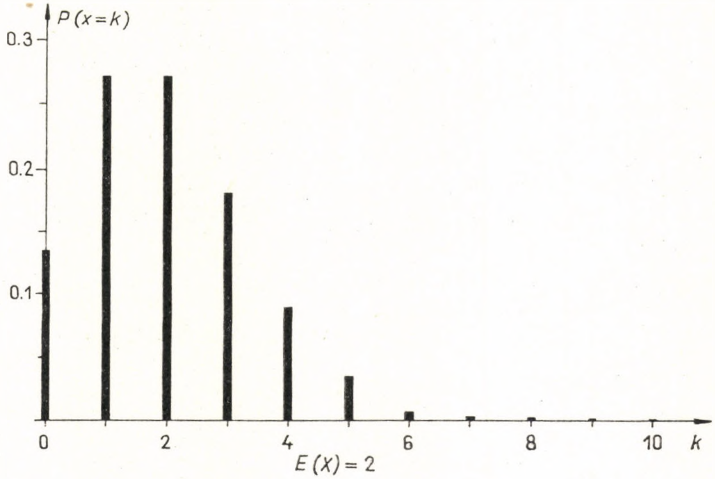


FIG. 10. Poisson distribution with parameter $\lambda = 2$

Values for the Poisson distribution are given in Table V (p. 380).

2.9.3 Binomial Distribution

The random variable $\varkappa$ is distributed according to the binomial law, with the constants n and p, if

$$P(\varkappa = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

Here n determines the size of the sample space (which is $n + 1$), while the parameter of the binomial distribution, p, may take its value from the interval $(0, 1)$. (Figure 11.)

Let us consider now n independent trials having possible outcomes A and $\bar{A}$ with probabilities p and $(1 - p)$ respectively.

(This is the situation in Experiment 1 if the choice is made with replacement and A denotes the event that the piece chosen is defective.)

We will prove this statement for the sake of simplicity in the special case $n = 4$, $k = 2$. In this case the event $\{\varkappa = 2\}$ may be obtained in $\binom{4}{2} = 6$ different ways, according to the possible succession of the events A and $\bar{A}$ in a trial:

$$\{\varkappa=2\}=AA\bar{A}\bar{A} + A\bar{A}A\bar{A} + A\bar{A}\bar{A}A + \bar{A}AA\bar{A} + \bar{A}A\bar{A}A + \bar{A}\bar{A}AA.$$

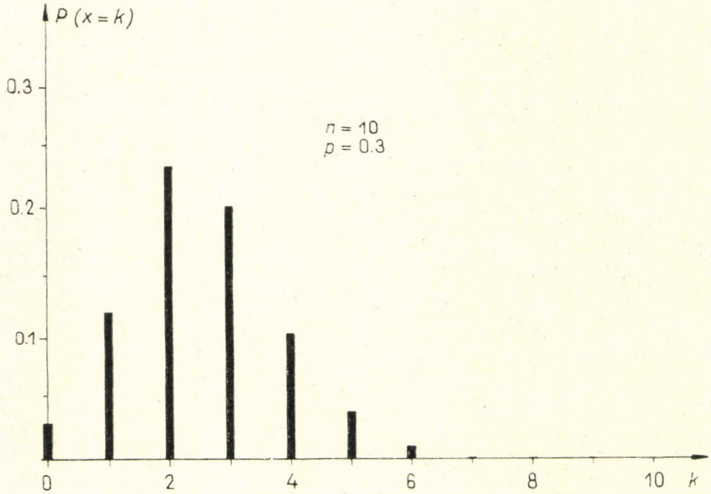


FIG. 11. Binomial distribution

These six events are disjoint, and because of the independence within each term, the probability of each of them equals $p^2(1-p)^2$. Consequently, $P(\varkappa = 2) = \binom{4}{2} p^2(1-p)^2$, which reflects the general case.

The first two moments and the variance are

$$E(\varkappa) = np, \qquad D^2(\varkappa) = np(1-p),$$

$$E(\varkappa^2) = D^2(\varkappa) + [E(\varkappa)]^2 = np(1-p) + n^2p^2.$$

The proof can be carried out with the aid of direct calculation by showing, for example, the validity of

$$E(\varkappa) = \sum_{k=0}^{n} k \binom{n}{k} p^k(1-p)^{n-k} = np.$$

for the first relation.

But there is another very simple, probabilistic way which, in many cases, leads to the result without any calculations. Let us make use now of the following random variables introduced earlier:

$$\eta_i = \begin{cases} 1, & \text{if } A \text{ occurs in the } i\text{th trial,} \\ 0, & \text{if } \bar{A} \text{ occurs in the } i\text{th trial.} \end{cases}$$

The η_i's are independent, with expected values

$$E(\eta_i) = 1 \cdot P(A) + 0 \cdot P(\bar{A}) = p$$

and variances

$$D^2(\eta_i) = (1-p)^2 \ \ P(A) + (0-p)^2 \ \ P(\bar{A}) = p(1-p).$$

According to the definition

$$\varkappa = \eta_1 + \eta_2 + \ldots + \eta_n$$

and by applying (III) and (VII′) from Section 2.6 we have

$$E(\varkappa) = E(\eta_1) + E(\eta_2) + \ldots + E(\eta_n) = np$$

$$D^2(\varkappa) = D^2(\eta_1) + D^2(\eta_2) + \ldots + D^2(\eta_n) = np(1-p).$$

The binomial distribution plays an important role in quality control, especially in acceptance sampling.

2.9.4 Exponential Distribution

The exponential distribution is determined by a single positive parameter $\lambda > 0$. The random variable X is distributed exponentially if its distribution function has the form

$$P(X \leqslant x) = 1 - e^{-\lambda x}, \qquad x \geqslant 0.$$

For the expected value and standard deviation we have

$$E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda}.$$

Under certain conditions the time between two consecutive failures in some machine equipment follows an exponential distribution. The distribution of the life span of single component parts differs in general, however, from the exponential distribution.

Therefore, in all cases arising in practice concerning the life span of single parts, a thorough statistical analysis is needed.

The exponentially distributed random variable has the following interesting property:

$$P(X \leqslant x + y) \mid X \geqslant y) = P(X \leqslant x) = 1 - e^{-\lambda x}.$$

This means that if equipment with exponentially distributed life span has already reached a certain age y without failure, then it has at this age the same chances of "survival" as at the beginning; it does not grow older but remains "eternally young."

2.9.5 Gamma Distribution

Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed random variables with the common exponential distribution $1 - e^{-\lambda x}$, $x > 0$. Then the distribution of their sum

$$Y_n = X_1 + X_2 + \ldots + X_n$$

is the *gamma distribution of order n with parameter λ*;

$$P(Y_n \leqslant x) = \int_0^x \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \, dt, \quad x \geqslant 0.$$

For the expected value and standard deviation we have

$$E(Y_n) = \frac{n}{\lambda} \quad D(Y_n) = \frac{\sqrt{n}}{\lambda}.$$

2.9.6 The Chi-Square (χ^2) Distribution

Let $X_1, X_2, \ldots, X_n$ be standard normal variables. Then the sum of their squares

$$\chi_n^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

has the so-called chi-square or χ^2 distribution belonging to the parameter* n. The distribution function is the following:

$$P(\chi_n^2 \leqslant x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} \int_0^x t^{n/2-1} e^{-t/2} \, dt, \quad x \geqslant 0.$$

---

* This parameter is often called the number of degrees of freedom.

For the expected value and standard deviation we have (Fig. 12)

$$E(\chi_n^2) = n,$$

$$D(\chi_n^2) = \sqrt{2n}\,.$$

If the X_i's are distributed according to $N(0, \sigma)$, then the random variable χ_n^2/σ^2 has the above χ^2 distribution. From this we may obtain the distribution function of χ_n^2; the latter will of course depend on σ.

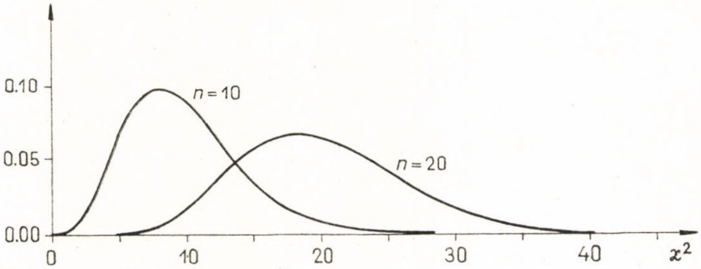


FIG. 12. Density functions of χ^2 distributions

One may observe that if $n = 2m$, then the χ^2 distribution becomes equivalent to the gamma distribution. Replacing the parameters n and λ of the gamma distribution by m and $\frac{1}{2}$ respectively, we obtain the χ^2 distribution with $n = 2m$.

Furthermore, it can be proved that if $\overline{X}$ denotes the arithmetic mean of the X_i given above, then the sum

$$\frac{(n-1)\,S^2}{\sigma^2} = \frac{(X_1-\overline{X})^2 + (X_2-\overline{X}) + \ldots + (X_n-\overline{X})^2}{\sigma^2}$$

is χ^2-distributed with parameter $n - 1$.

2.9.7 The Weibull Distribution

For the distribution of the life span of single component parts the so-called Weibull distribution is sometimes a good approximation. The random variable X with Weibull distribution has a distribution function with the following form:

$$P(X \leqslant x) = 1 - e^{-ax^b},\ x \geqslant 0,\quad a > 0,\quad b > 0.$$

The expected value of this random variable equals

$$E(x) = \frac{1}{a^{1/b}}\,\Gamma\left(\frac{1}{b}+1\right),$$

while its variance is

$$D^2(X) = \frac{1}{a^{2/b}}\left[\Gamma\left(\frac{2}{b}+1\right) - \Gamma^2\left(\frac{1}{b}+1\right)\right].$$

2.9.8 Student or *t* Distribution

Let $Y, X_1, X_2, \ldots, X_n$ be independent random variables with the common distribution $N(0, \sigma)$. Then the random variable

$$t_n = \frac{Y}{\sqrt{(X_1^2 + X_2^2 + \ldots + X_n^2)}} = \frac{Y}{\chi_n}$$

follows the so-called Student or t distribution with parameter n. The distribution function has the form

$$P(t_n < x) = \frac{1}{\sqrt{n\pi}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\int\limits_{-\infty}^{x}\frac{dt}{\left[1+(t^2/n)\right]^{(n+1)/2}}, \quad -\infty < x < \infty.$$

The shape of the density function of the random variable (Fig. 13) is similar to that of the Gaussian density. Indeed, if $n \to \infty$, then t_n

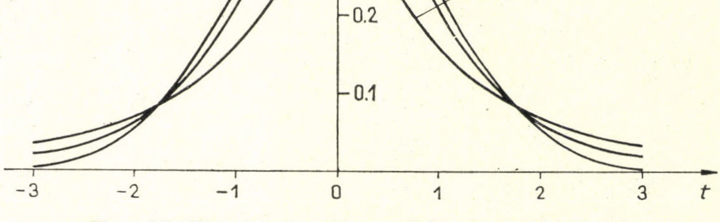


FIG. 13. Density functions of Student distributions

converges in distribution to the normal law. This means that if n is large enough (e.g., $n > 80$) then the following relation holds to a good approximation:

$$P(t_n < x) \approx \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2}\, dt, \qquad -\infty < x < \infty .$$

For $n = 1$ the random variable $t_1 = Y/\,|X_1|$ has the distribution function

$$P(t_1 < x) = \int_{-\infty}^{x} \frac{dt}{1+t^2}, \qquad -\infty < x < \infty .$$

This is the so-called Cauchy distribution. Since the expected value of $|t_1|$ is infinite,

$$E(|t_1|) = \int_{-\infty}^{\infty} \frac{|t|\, dt}{1+t^2} = \infty ,$$

the expected value of t_1 does not exist. For t_n only the first $(n-1)$ moments exist; the nth does not.

For the expected value and variance of t_n in the case of $n \geqslant 2$ we have

$$E(t_n) = 0 ,$$

$$D^2(t_n) = \frac{n}{n-2}$$

2.9.9 The Beta Distribution

The random variable β has a *beta distribution of first kind* with the pair of parameters (n_1, n_2) if its distribution function is

$$P(\beta \leqslant x) = \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \int_{0}^{x} t^{n_1-1}(1-t)^{n_2-1}\, dt, \quad 0 \leqslant x \leqslant 1 .$$

The possible values of β range over the finite interval (0, 1). For the expected value and variance we have

$$E(\beta) = \frac{n_1}{n_1+n_2}$$

and

$$D^2(\beta) = \frac{n_1 n_2}{(n_1+n_2)^2\,(n_1+n_2-1)}$$

Transforming the variable β into $\beta' = \beta/(1 - \beta)$, we obtain a random variable that takes on values in the half-axis $(0, \infty)$. Its distribution is the so-called *beta distribution of the second kind* with the pair of parameters (n_1, n_2):

$$P(\beta' \leqslant x) = \frac{\Gamma(n_1 + n_2)}{\Gamma(n_1)\Gamma(n_2)} \int\limits_0^x \frac{t^{n_1 - 1}}{(1 + t)^{n_1 + n_2}}\, dt, \quad 0 < x < \infty .$$

It can be shown that if $X_1, X_2, \ldots, X_{f_1}$ and $Y_1, Y_2, \ldots, Y_{f_2}$ are independent random variables distributed according to $N(0, 1)$, then the random variable

$$\tau = \frac{X_1^2 + X_2^2 + \ldots + X_{f_1}^2}{Y_1^2 + Y_2^2 + \ldots + Y_{f_2}^2}$$

has beta distribution of the second kind with the pair of parameters $(f_1/2,\ f_2/2)$.

For the expected value and variance of τ we have

$$E(\tau) = \frac{f_2}{f_2 - 2} \qquad \text{for} \quad f_2 > 2 ,$$

$$D^2(\tau) = \frac{2 f_1(f_1 + f_2 - 2)}{(f_2 - 2)^2 (f_2 - 4)} \qquad \text{for} \quad f_2 > 4 .$$

2.9.10 The Logarithmic Normal Distribution

Let Y be a nonnegative random variable whose logarithm $X = \ln Y$ is distributed according to $N(\ln a, \sigma)$. Then we say that Y has a logarithmic normal distribution. Since

$$P(\ln Y \leqslant x) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{x} \exp\left(- \frac{(t - \ln a)^2}{2\sigma^2}\right) dt ,$$

we obtain, by the substitution $x = \ln y$, the distribution of Y:

$$P(Y \leqslant y) = L(y) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\ln y} \exp\left(- \frac{(t - \ln a)^2}{2\sigma^2}\right) dt, \quad 0 < y < \infty .$$

For the density function we have

$$L'(y) = l(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(\frac{(\ln y - \ln a)^2}{2\sigma^2}\right), \quad 0 < y < \infty ,$$

for which

$$l(y) = \frac{1}{\sigma y} \varphi \left(\frac{\ln y - \ln a}{\sigma} \right), \quad 0 < y < \infty,$$

where $\varphi(x)$ denotes as usual, the density function of the standard normal variable.

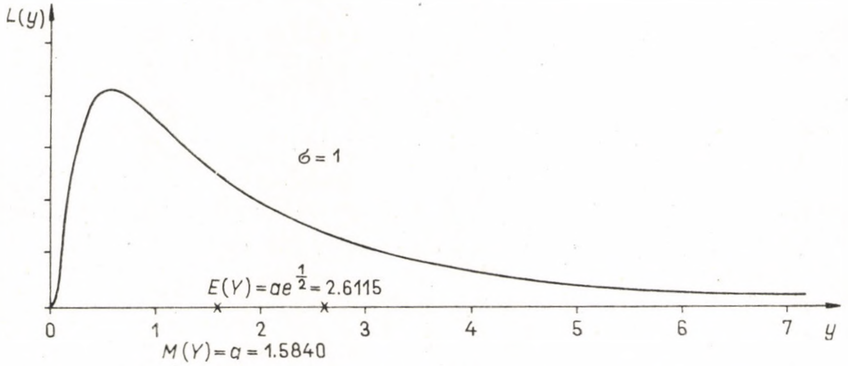


FIG. 14. Relative position of the expected value and of the median (in the case of the logarithmic normal distribution)

The expected value and variance of ln Y are, on account of the property of the normal distribution,

$$E(\ln Y) = \ln a,$$

$$D^2(\ln Y) = \sigma^2.$$

However, for the logarithmic normal variable Y the following is valid (see Fig. 14):

$$E(Y) = ae^{\sigma/2}$$

$$D^2(Y) = a^2 e^{\sigma^2}(e^{\sigma^2} - 1).$$

The median of Y equals a:

$$L(a) = \frac{1}{2}.$$

Logarithmic normal distributions appear in milling processes, where often the size (diameter) of the small pieces is distributed according to this law.

2.9.11 The Uniform Distribution

The random variable Y taking on values in the interval $(0, 1)$ is distributed uniformly if its distribution function has the form

$$U(x) = P(Y \leqslant x) = \begin{cases} 0, & \text{if } x \leqslant 0, \\ x, & \text{if } 0 \leqslant x \leqslant 1, \\ 1, & \text{if } 1 \leqslant x. \end{cases}$$

The density function is

$$U'(x) = u(x) = \begin{cases} 1, & \text{if } 0 \leqslant x \leqslant 1, \\ 0, & \text{otherwise.} \end{cases}$$

The random variable Y is uniformly distributed in the interval (a, b) $(a < b)$ if its distribution function has the form

$$U_{a,b}(x) = P(Y \leqslant x) = \begin{cases} 0, & \text{if } x < a, \\ \dfrac{x - a}{b - a}, & \text{if } a \leqslant x \leqslant b, \\ 1, & \text{if } x > b. \end{cases}$$

The density function of Y is

$$u_{a,b}(x) = \begin{cases} \dfrac{1}{b - a}, & \text{if } a \leqslant x \leqslant b, \\ 0, & \text{otherwise}. \end{cases}$$

Let X be a continuous random variable with the strictly increasing distribution function $F(x)$. Replacing the functional variable x by X in the distribution function we obtain a random variable $Y = F(X)$ distributed uniformly in $(0, 1)$. The simple proof of this fact is the following, where $0 \leqslant y \leqslant 1$:

$$P(Y \leqslant y) = P[F(X) \leqslant y] = P[X \leqslant F^{-1}(y)] = P[F^{-1}(y)] = y.$$

2.9.12 Hypergeometric Distribution

Let us have a lot of products of size N containing M defective pieces and $N - M$ good ones. Taking a sample of size n out of the lot randomly, let us denote the number of defective items in the sample by $\varkappa$. Random sampling in this case means that the n pieces are taken simultaneously from the lot; in other words, the sample elements are selected one after the other *without replacement*. It can easily be seen that the random variable $\varkappa$ is distributed according to the following law:

$$P(\varkappa = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad k = 0,1,2,\ldots, \min\ (n, M),$$

which is the so-called *hypergeometric distribution.*

The first moment and the variance, with the notation $p = M/N$, are

$$E(\varkappa) = n\frac{M}{N} = np, \quad D^2(\varkappa) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1} = np(1-p)\frac{N-n}{N-1}\ .$$

Since $(N - n)/(N - 1) < 1$, the variance of the number of defectives $\varkappa$ in the case of sampling without replacement is always smaller than in the case of sampling *with* replacement, the variance of the binomial distribution being $np(1 - p)$ (see Section 2.9.3).

Although in the practice of quality control the sampling is done without replacement, the binomial distribution is used. This can be justified heuristically with the argument given in Section 1.7: If a single piece is selected, then after replacement the probability of selecting the same piece on the next is about $1/N$, which is a very small quantity. But, continuing this idea, we can arrive at a mathematical theorem that holds when $n \ll N$ (e.g., $n < N/10$), the two probabilities are near each other:

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \sim \binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}.$$

This justifies our practice of using the binomial model almost throughout, that is, of considering the sampling as if it were performed with replacement.

2.10 THE CENTRAL LIMIT THEOREM

This limit distribution theorem has been given the name "central" because of its great significance.

Let $X_1, X_2, \ldots$ be an infinite sequence of independent identically distributed random variables with common expected value E and variance D^2; then the central limit theorem states that the distribution of the standardized form of the sum $Y_n = X_1 + X_2 + \ldots + X_n$ converges to the distribution function of the standard normal variable:

$$\lim_{n\to\infty} P\left(\frac{Y_n - E(Y_n)}{D(Y_n)} \leqslant x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt\,,$$

where $E(Y_n) = nE$ and $D(Y_n) = \sqrt{n}D$.

This and related theorems explain the fact that a Gaussian distribution appears in cases where the outcome of the experiment depends on many factors, the effects of which are individually small and independent, but cumulative. In all practical cases, however, it is advisable to be careful, since it may be that the necessary requirements are not fulfilled and thus the resulting limiting distribution differs from the normal law. (If, for instance, the resulting effect is not the sum of the individual small effects but their product, then the limiting distribution is of logarithmic normal type.) For this reason, in the actual experiment (production process) the distribution of the variable has to be examined in order to determine whether it can be treated as Gaussian distribution or not. On the other hand, even if the distribution of the single variables differs from that of the normal law, their sum may be normally distributed if the conditions of the corresponding limit theorems are satisfied.

2.11 BAYES' THEOREM FOR CONTINUOUS VARIABLES

Let X be a continuous random variable with (prior) distribution $P(X \leqslant x) = F(x) = \int_{-\infty}^{x} f(t)\,dt$. Let η be a continuous random variable as well. Then, according to Bayes' theorem, the conditional density function of X under the condition that an observation on Y resulted in y, i.e., the posterior density of X, is given by

$$f(x \mid y) = \frac{f(y \mid x)\, f(x)}{\int_{-\infty}^{\infty} f(y \mid t)\; f(t)\,dt}\,, \quad -\infty < x < \infty. \tag{10}$$

Here $f(x \mid y)$ and $f(y \mid x)$ denote the conditional densities of X and Y under the condition $Y = y$ and $X = x$, respectively.

Conditional densities can be interpreted by starting from discrete distributions and using a limiting process. The same applies for conditional probabilities given a condition of probability 0. For a more rigorous treatment, see [A 17 or A 30].

Using the notation

$$P(X < x \mid Y = y) = F(x \mid y),$$

for the conditional (posterior) distribution, the theorem can be written in the form

$$F(x \mid y) = \frac{\int_{-\infty}^{x} f(y \mid t)\, f(t)\, dt}{\int_{-\infty}^{\infty} f(y \mid t)\, f(t)\, dt}, \quad -\infty < x < \infty. \tag{10'}$$

Integration, with respect to x in Eq. (10) and substitution of $x = \infty$ in Eq. (10′) results in unity; i.e., we have distribution functions.

As in the discrete case (Section 1.9) these formulas may be easily obtained from the formulae on conditional distributions and densities given in Section 2.2.

The case when one of the variables is continuous and the other is discrete does not cause any difficulties. We will see this in Section 4.2.4.4 where the example given in Section 1.9 will be solved by applying the *Bayesian method*.

3 FUNDAMENTALS OF SAMPLING

3.1 THE SAMPLE

3.1.1 Definition of the Sample

We have already met the concept of the random variable. Information about the result of an observation can be obtained from its distribution function. Moreover, certain theoretical quantities derived from the distribution function help us to get a good summary of its behavior. In general, however, we do not know the distribution and its characteristics in advance. We wish to determine these theoretical quantities. What we can do is to make measurements on the random variable. *A given number, say n, of independent measurements on a random quantity X will be called a sample on X of size n.* With the aid of the sample we will be able to approximate statistically; i.e., to estimate the true theoretical characteristics. Furthermore, it is often of interest to decide questions of the following type: Can certain assumptions concerning the theoretical characteristics be accepted as valid, or are they to be rejected as wrong?

First of all, we shall consider the sample itself. Here we must call attention to a certain duality. From the point of view of the theory of mathematical statistics we speak about "n observations on X" without actually carrying out the measurements; i.e., a sample of size n is an n-tuple of independent random variables $X_1, X_2, \ldots, X_n$, where X_1 is defined as the first observation on X and so on. On the other hand, the name "sample" will also be used for a realization of the n measurements. Assuming that we have carried out the observations on X and obtained the values $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$, we say that the sampling has resulted in the "sample" $x_1, x_2, \ldots, x_n$. This will not lead to any confusion.

We must emphasize that according to the definition of a sample, its elements X_i are observations on the *same* random quantity, i.e., they are not only *independent* but also *identically distributed* random variables:

$$P(x_i \leqslant x) = F(x), \; i = 1, 2, \ldots, n;$$

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F(x_i),$$

where $F(x)$ denotes the common distribution function of the X_i's, while $F(x_1, x_2, \ldots, x_n)$ is their joint distribution function.

The fact that the order of the sample elements was defined in accordance with the order of the observations is closely connected with the problem of independence. In some practical problems, such sampling leads to a so-called ordered sample, in which the sample elements occur in the order of their magnitudes; such elements are not independent. This situation arises in cases of testing for life span (bulbs, resistors, etc.), where all sample elements are subjected to loading at the same time. To get independent random variables we have to decide before the experimentation which piece will be regarded as the first sample element, which one as the second, etc.

If sampling consists of selecting from a lot, i.e., from a finite population, then we always assume that the sampling occurs with replacement; this leads to independent sample elements. In statistical literature the sample space itself is often called the "population", from which the sample elements are "chosen".

3.1.2 Sample Size

According to the laws of large numbers, the relative frequency tends stochastically toward the probability. With increasing number of observations the arithmetic mean of identically distributed random variables converges stochastically to the expected value of the random variable. This means that for an "exact" determination of these theoretical quantities, an infinite, or at least a very large, number of observations is required.

In quality control we often need good information on certain distribution functions and their characteristics. For this purpose we have to take a *large sample;* what a large sample really means depends on various circumstances, such as, for example, the size of the lot which is to be examined. In connection with control during production a preliminary sample is taken, usually consisting of 300–500 items.

On the other hand, the aim is sometimes no more than a comparison or the checking of a variable or the obtaining of some complementary information, etc. In such cases we are satisfied with *small samples.* A small sample may be of size 4–10 or 30–40, depending on the problem and on the statistical method used.

These sample sizes are taken from practical experience where economic considerations are also involved. Exact methods for the determination of the sample size in different situations will be given later. In certain cases (e.g., in scientific research work) the necessary sample size may be as large as several thousand.

We have tacitly assumed until now that the sample size is fixed before experimentation. On the other hand, there are sampling methods in which the sample size is not fixed in advance, but is a random variable as well. This is the case, for example, with the so-called double and multiple sampling plans.

In the double procedure we start with a first sample of given size n_1; depending on its result we either make a final decision or take a second sample of size n_2 and then base the final decision on the result of the combined sample. In this case the sample size is a random variable ν, which may take on the values n_1 or $n_1 + n_2$. From the practical point of view the average number of observations, i.e., the expected value of ν, $E(\nu)$, is of great interest; if this value is smaller than the sample size of a statistically equivalent procedure based on a simple sample, then double sampling may be more advantageous. Whether it really is, depends, of course, on the practical consequences of taking a second sample. In examining the breaking strength of concrete, one experiment sometimes requires 25–30 days, and thus it is not advisable to take a second sample.

On the other hand, if the procedure needed for obtaining a sample element does not take long, then even so-called sequential sampling may be feasible. This method, elaborated by A. Wald, does not restrict the sample size at all: It may be infinite, at least theoretically. Nevertheless, the expected value of the sample size is in certain cases only 40–60% of the corresponding size of a simple sample. We shall return to multiple and sequential sampling in Chapter III, Section 2.2.

3.2 SAMPLE CHARACTERISTICS; THE STATISTIC

3.2.1 Working Out the Statistical Data

We turn again to Experiment 2 in which cylinders are produced by a given machine, the nominal value of the length X of the cylinder being 23 mm. We wish to take a sample of "medium" size on X, i.e., a sample of size $n = 120$. This is carried out by measuring the length of 120 cylinders as they come out of the machine. If our measuring instrument is able to detect very small differences, the 120 values will probably be different from each other. But we are satisfied now with an accuracy of four figures; this means that all actual values between the limits 23.015 and 23.025 will be registered as 23.02. Thus the values obtained will not all be different.

Let us denote by $(X_1, X_2, \ldots, X_{120})$ the sample before carrying out the measurements and by $(x_1, x_2, \ldots, x_{120})$ the observed values that are given in Table 5; e.g., $x_1 = 22.97$, $x_{10} = 22.98$, $x_{30} = 22.89$. If the tolerance limits are 22.88 and 23.12, then we may wish to know the number of products that are defective with respect to length. This leads to the fraction defective, which is the relative frequency of the event $\{X < 22.88 \text{ or } X > 23.12\}$ and which gives us certain information about the probability of the event that an item will turn out to be defective in this production process.

We turn now to a further investigation of our data. We are interested in how large the average deviations from the prescribed value are and how often we get relatively large deviations, even if these are within the tolerance limits.

Roughly speaking, we will consider the following: the value around which the sample elements cluster; how and to what extent they disperse; and how their properties can be characterized by means of suitably chosen quantities.

First of all, we have to arrange the data according to their order of magnitude. But because the number of different figures is nearly 30, this arrangement does not give us a clear picture of the regularities. For this reason the sample elements need to be further grouped into

TABLE 5

LENGTH OF CYLINDERS IN MILLIMETERS

22.97	23.00	22.90	22.98
23.05	23.04	23.00	23.02
22.98	23.03	23.04	23.04
23.04	23.00	22.91	22.99
23.10	22.88	23.00	22.91
23.05	23.00	22.95	22.96
22.95	22.96	22.96	22.94
22.98	22.90	23.01	22.98
23.01	23.00	22.82	22.93
22.98	23.01	23.05	23.05
23.02	22.99	23.02	22.99
22.92	22.91	23.07	22.94
22.92	22.99	23.09	23.14
23.11	22.92	22.94	22.99
23.03	23.08	23.07	22.97
23.06	22.91	23.06	23.03
22.97	23.08	22.96	23.06
23.09	23.07	22.93	22.89
22.95	23.05	22.97	23.07
23.07	23.95	23.01	22.96
22.90	23.02	23.03	22.96
22.94	22.87	23.04	22.95
22.98	22.99	22.97	23.01
23.03	23.02	22.98	23.00
22.97	22.99	22.95	22.98
22.95	23.07	22.96	22.97
23.01	22.99	23.12	22.94
22.99	22.95	23.00	22.94
22.88	23.03	22.89	23.01
22.89	23.03	22.93	23.06

cells. As indicated in Table 6, we have chosen 12 cells, the respective intervals given in the first column. (For example, the third cell corresponds to the event $\{22.875 < X < 22.905\}$.) The second column contains the midpoints of the cells. The third column contains the frequencies, and the fourth the relative frequencies of the events defined by the cells.

Denoting by a_{i-1} and a_i the endpoints of the ith cell, for sufficiently large values of n the relative frequency of a cell approximates its probability:

$$\frac{f_i}{n} \approx P(a_{i-1} < X < a_i) = F(a_i) - F(a_{i-1}), \quad i = 1, 2, \dots .$$

TABLE 6

Cells	Midpoints	Frequencies	Relative frequencies
22.815–22.845	22.83	1	0.0083
22.845–22.875	22.86	1	0.0083
22.875–22.905	22.89	8	0.0666
22.905–22.935	22.92	10	0.0833
22.935–22.965	22.95	21	0.1750
22.965–22.995	22.98	24	0.2000
22.995–23.025	23.01	21	0.1750
23.025–23.055	23.04	17	0.1417
23.055–23.085	23.07	11	0.0916
23.085–23.115	23.10	4	0.0333
23.115–23.145	23.13	2	0.0167
23.145–23.175	23.16	0	0.0000

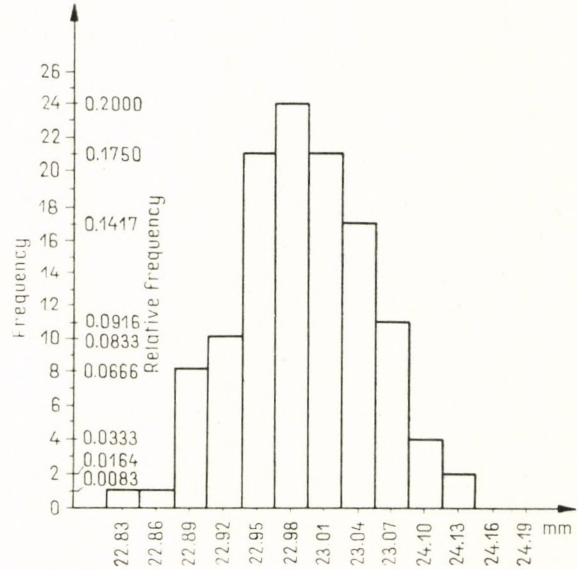


FIG. 15. Graphical representation of the data given in Table 2

One may observe that in the case of cells of equal length the cell frequencies show a tendency to increase when starting from either end and approaching the middle. This is clear from Fig. 15, in which a rectangle represents each cell, its area being proportional to the frequency or to the relative frequency of the cell. The diagram obtained is called a frequency histogram.

In most practical cases 8–15 cells prove to be convenient. The cells are not necessarily of the same length.

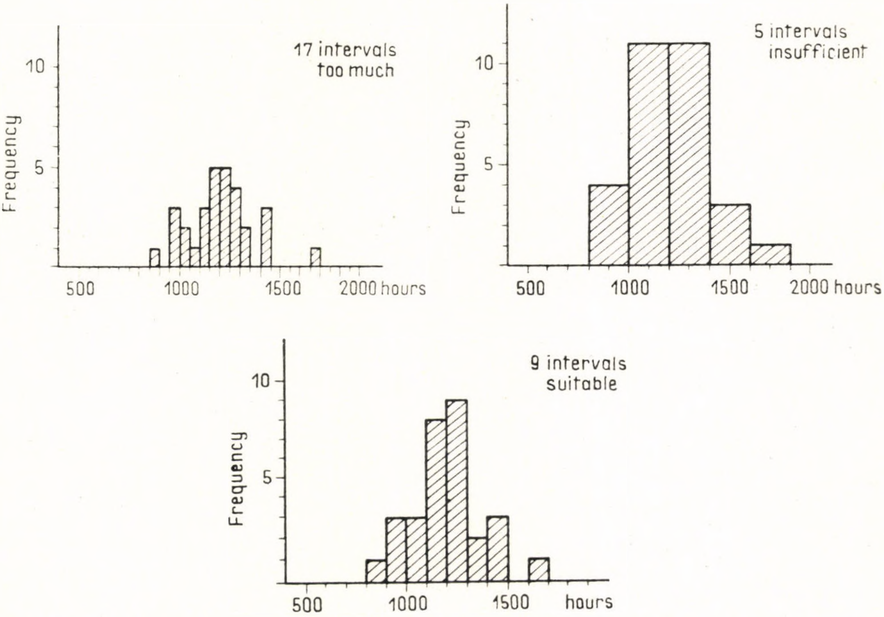


Fig. 16. Graphical representation of the same statistical data using different lengths for the intervals (cells)

The choice of the number and length of the cells is of greater importance if the sample size is relatively small. In this case the character of the distribution is obscured if too many or too few cells are used. This fact is illustrated in Fig. 16, where the different histograms are constructed from the same sample of size 30.

There is, however, no general rule regarding the appropriate number and length of the cells. A suitable choice depends on the problem being treated and on the statistical method used. Very few theoretical results exist in this respect, but practical experience helps in finding good solutions in different situations.

3.2.2 The Most Frequently Used Sample Characteristics

The frequency diagram gives us certain information on the distribution of the random variable: it shows an interval containing most of the distribution, etc. Our next aim is to construct some quantities that allow a concise characterization of the distribution when knowledge of the true distribution function is lacking. In other words, we construct sample functions for estimating the theoretical characteristics. The nature of this estimation will be considered in Section 4.2.

The two main types of characteristics are the measures of location and the measures of dispersion.

A sample characteristic is always a function of sample elements. Having a sample $(X_1, X_2, \ldots, X_n)$, any function $\alpha(X_1, X_2, \ldots, X_n)$ of the sample elements that has as a random variable in a well-defined distribution will be called a *statistic*.

We now present some useful statistics:

3.2.2.1 The Sample Mean or Arithmetic Mean. The sample mean is the arithmetic mean of the sample elements; we shall denote it throughout by $\overline{X}$:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n};$$

that this notation does not explicitly indicate the sample size n will not cause confusion.

The sample elements X_i differ from the sample mean by $X_i - \overline{X}$; it is easy to see that

$$\sum_{i=1}^{n} (X_i - \overline{X}) = 0$$

is always true.

The sample of size $n = 120$ given in Table 5 has the sample mean

$$\overline{x} = \frac{22.97 + 23.05 + \ldots + 22.94 + 23.01}{120} = 22.9897.$$

If less accuracy is needed, we can compute an approximation to $\overline{x}$ from Table 6. In this case we regard all values lying in the same cell as identical and substitute for them the midpoint of the cell, indicated in the second column. Using the notation y_i as the midpoint of the ith cell and f_i as the corresponding frequency, the approximate value of the sample mean takes the form

$$\tilde{x} = \frac{1}{n} \sum_{i=1}^{r} f_i y_i,$$

where r denotes the number of cells. In our example we have $r = 12$; carrying out the computation we obtain

$$\tilde{x} = 22.9900.$$

Thus the difference from the exact value of the sample mean is less than 0.25%. This difference is not always so small, but if the cells are not too wide and/or the values are situated approximately around the cell midpoint then the error is negligible.

For the arithmetic mean the following simple properties hold:

(a) If the values $x_1, x_2, \ldots, x_n$ have arithmetic mean $\bar{x}$, then the arithmetic mean of the values $x_1 + a, x_2 + a, \ldots, x_n + a$ is $\bar{x} + a$. This gives for $a = -\bar{x}$ the relation

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) = \bar{x} - \bar{x} = 0.$$

(b) If the arithmetic mean of $x_1, x_2, \ldots, x_n$ is $\bar{x}$, then the values $ax_1, ax_2, \ldots, ax_n$ have the arithmetic mean $\overline{ax}$.

(c) If $\bar{x}$ and $\bar{y}$ are the arithmetic means of the values $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$, respectively, then $x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n$ have the arithmetic mean $\bar{x} + \bar{y}$.

Finally, we wish to clarify the relationship between the sample mean and the expected value of the sample elements. We shall refer to the approximate expression $\tilde{x}$ introduced above. Let us denote by $-\infty = a_0 < a_1 < a_2 < \ldots < a_{r-1} < a_r = \infty$ the endpoints of the cells, where we extend the first and the last cell to infinity. Then the cells correspond to the events

$$A_i = \{a_{i-1} < X < a_i\}, \quad i = 1, 2, \ldots, r,$$

and f_i/n is the relative frequency of the event in the sequence of observations yielding the sample. Denoting by $f(x)$ the common density function of the sample elements, for the probability of A_i we have

$$P(A_i) = P(a_{i-1} < X < a_i) = \int_{a_{i-1}}^{a_i} f(y)\, dy.$$

Let us assume now that n is sufficiently large and thus the relative frequency is close to the probability. Then we may write

$$\tilde{x} = \frac{1}{n} \sum_{i=1}^{r} f_i y_i = \sum_{i=1}^{r} y_i \frac{f_i}{n} \approx \sum_{i=1}^{r} y_i \int_{a_{i-1}}^{a_i} f(y)\, dy.$$

According to the mean-value theorem of the integral calculus, the following relation holds:

$$y_i \int_{a_{i-1}}^{a_i} f(y)\,dy \approx \int_{a_{i-1}}^{a_i} y\,f(y)\,dy;$$

the sum of the right-hand terms gives the expected value of the sample elements, from which we have

$$\tilde{x} \approx E(X).$$

This means that the sample mean fluctuates around the expected value. The fact that it tends stochastically to the expected value was mentioned in the section on the laws of large numbers.

3.2.2.2 The Order Statistics. Let us arrange the elements of a sample of size n according to their order of magnitude; let $X_1, X_2, \ldots, X_n$ be the independent sample elements, and let us denote by X_1^* the smallest among them; the smallest of the remaining $(n - 1)$ elements will be denoted by X_2^*; and so on. If the *common distribution function $F(x)$ of the sample elements is continuous*, then the probability that two or more among the X_i's will coincide equals zero; so they can be taken as being different. Thus with probability 1 we may write

$$X_1^* < X_2^* < \ldots < X_n^*.$$

X_k^* is called the kth order statistic.

At the end of this chapter, in Section 3.3 we will consider the fundamentals of the theory of order statistics. There the procedure of ordering will be explained and the distribution of the kth order statistic will be determined. The theory of order statistics has long been applied with success in quality control. Many methods have been elaborated for the use of samples with ordered elements; such methods have the great advantage that they do not involve any computations, merely the listing of the observed data. Upon marking the place of each sample element on the real line, the elements automatically become ordered.

Order statistics are statistics as defined above: they depend on the sample elements. By means of them we are able to construct further sample characteristics that prove to be very useful in practice.

3.2.2.3 The Sample Median. The sample median is the middle term in the ordered sample if n is odd, and it is the arithmetic mean of the two middle terms if n is even; i.e., for $n = 2m + 1$ the sample median is

X^*_{m+1}, while for $n = 2m$, it is $(X^*_m + X^*_{m+1})/2$. In our previous example, $n = 120$ and $X^*_{60} = X^*_{61} = 22.99$; the sample median is also 22.99.

The sample median is generally different from the sample mean; its determination is simpler than that of the sample mean. On the other hand, the sample mean is a more informative and sensitive characteristic; changing the value of one sample element may leave the median unchanged, but the sample mean varies if a sample element is replaced by another value.

Sample mean and sample median have important roles in statistical quality control; the former because of its theoretical significance, the latter because it can be determined very easily, especially for small samples.

3.2.2.4 The Mean Deviation of the Sample. Let $\bar{X}$ be the sample mean belonging to the sample $X_1, X_2, \ldots, X_n$. The mean deviation d of the sample is the arithmetic mean of the absolute values of the deviations $X_i - \bar{X}$:

$$d = \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}| = \frac{|X_1 - \bar{X}| + |X_2 - \bar{X}| + \ldots + |X_n - \bar{X}|}{n}.$$

The mean deviation itself is not often used in practice. In the case of a positive variable the so-called relative mean deviation

$$d_r = \frac{d}{\bar{X}}$$

is more often applied, mainly in the textile industry.

The computation of d may be carried out in the following way: Denoting by k the index for which $X^*_k < \bar{X} < X^*_{k+1}$, then

$$d = \frac{2\sum_{i=1}^{k}(\bar{X} - X^*_i)}{n} = \frac{2\sum_{i=k+1}^{n}(X^*_i - \bar{X})}{n}.$$

The validity of this relation can be easily seen if we take into account the known relation $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$, which can be written in the form $\sum_{i=1}^{n}(X^*_i - \bar{X}) = 0$, and from which the equality $\sum_{i=1}^{k}(\bar{X} - X^*_i) = \sum_{i=k+1}^{n}(X^*_i - \bar{X})$ follows. This means the sum of the absolute values of the negative deviations is equal to the sum of the positive deviations, which justifies the computational method given above.

3.2.2.5 The Sample Variance and the Standard Deviation. The sample variance is also derived from the deviations $X_i - \bar{X}$ of the individual sample elements, and is defined by the formula

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}.$$

The sample variance, which we shall call simply the variance if there is no danger of misunderstanding, reflects the scattering of the values around the expected value. Its square root is the so-called *standard deviation*:

$$s = \left(\frac{(X_i - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1} \right)^{1/2}.$$

The introduction of the divisor $(n - 1)$ instead of n, which would give the arithmetic mean of the quadratic deviations, is due to theoretical considerations: as we will see in Section 4.2.2.1 this gives a better estimation of theoretical variance.

Using grouped data we get an approximate expression for the sample variance; with the notation used in the case of the sample mean, this is of the form

$$s^2 \approx \tilde{s}^2 = \frac{f_1(y_1 - \tilde{x})^2 + f_2(y_2 - \tilde{x})^2 + \ldots + f_r(y_r - \tilde{x})^2}{n-1}.$$

For large sample sizes we obtain the approximate relation

$$s^2 \approx \sigma^2 = \int_{-\infty}^{\infty} [y - E(X)]^2 f(y)\, dy.$$

This can be verified in a manner similar to the case of $\tilde{x}$. This means that if we take samples of size n repeatedly and compute the sample variances, these will fluctuate around the variance of the distribution.

If the groupings are carried out for cells of equal lengths, then the so-called *Sheppard correction* diminishes the error caused by grouping; denoting the length of the cells by h, the corrected form of the sample variance computed from grouped data is

$$\tilde{s}_h^2 = \tilde{s}^2 - \frac{1}{12} h^2.$$

The computation of the sample variance requires many numerical calculations. This is the case in particular when $\bar{x}$ has many figures. The calculation can be simplified with the aid of the following often used identity, the so-called Steiner relation known from the mechanics of rigid bodies:

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - A)^2 - n(A - \bar{X})^2$$

for any value of A. In our computations we shall choose for A a suitable value; e.g., an integer near to $\bar{X}$, or zero if the X_i's are small integers, etc.

In our numerical sample where we had $\bar{x} = 22.99$, the choice $A = 23.00$ will be suitable. Using the grouped data (Table 2) the calculation leads to the value $\tilde{s}^2 = 0.4163$; or, applying the Sheppard correction, $\tilde{s}^2 = 0.4162$, since $\frac{1}{12}\, h^2 = \frac{1}{12} \cdot 0.03^2 = 0.000075$.

The sample variance has the following simple properties:

(a) Adding a constant value a to all sample elements leaves the sample variance s^2 unchanged; indeed, this does not influence the relative position of the sample elements to each other and to the sample mean.

(b) If s^2 is the variance of the sample elements $X_1, X_2, \ldots, X_n$, then $|a|^2 s^2$ is the sample variance of the sample $aX_1, aX_2, \ldots, aX_n$.

(c) From the Steiner relation given above, the inequality

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \leqslant \frac{\sum_{i=1}^{n} (X_i - A)^2}{n-1} = d^2$$

follows, since we have increased the right side by the nonnegative quantity $n(A - \bar{X})^2$. This means that the sample variance for which $A = \bar{X}$ is the smallest among all possible quadratic mean deviations d^2 formed by any value A.

3.2.2.6 The Range. Whereas the mean deviation and the variance of the sample measure scattering on the basis of *deviations of the sample elements from the sample mean*, the *range* is a measure of scattering that takes into account only the length of the interval in which sample elements occur. The range, denoted by R, is defined as

$$R = X_n^* - X_1^*;$$

i.e., the range is the difference of the largest and smallest sample elements. Comparing the range with the sample variance, we may draw the same conclusions as we did at the comparison of sample mean and median: The change of the value of a single sample element will change the value of the variance, but not necessarily the value of the range. Computation of the range is very simple, however; consequently, it is of great importance in various applications, including statistical quality control.

3.2.2.7 The Coefficient of Variation. In many cases it is not the scattering of the values itself that we need, but its magnitude relative to the measure of location. If X is a positive random variable, then this relation is expressed as the ratio of the standard deviation and sample mean:

$$V = \frac{S}{\overline{X}}.$$

3.2.3 The Empirical Distribution

As we already know, the complete information on a random variable is contained in its distribution function, or equivalently in its density function if it is a continuous random variable. We now construct the empirical functions, which, in the absence of the distribution function or density function, give us certain information on the behavior of the random variable. As pointed out in connection with the sample characteristics, the nature of this approximation will be discussed in our further investigations on statistical estimation.

3.2.3.1 The Density Histogram. Earlier we defined the frequency or relative frequency diagram constructed from a sample. We shall now bring this diagram closer to the theoretical frequency function. Let us take, on a continuous random variable X, a sample of size n with sample elements $X_1, X_2, \ldots, X_n$. We group the data according to the cells (a_{i-1}, a_i), $i = 1, 2, \ldots, r$; here a_0 and a_r are near to the smallest and largest sample elements, respectively, with $a_0 < X_1^*$, $a_r > X_r^*$. We define the empirical density functions $f_n(x)$ as $g_i/(a_i - a_{i-1})$ on the ith cell, where g_i denotes the relative frequency of the cell. This means that

$$f_n(x) = \begin{cases} 0 & \text{if } x < a_0, \\ \dfrac{g_i}{a_i - a_{i-1}} & \text{if } a_{i-1} < x < a_i, \quad i = 1, 2, \ldots r, \\ 0 & \text{if } a_r < x. \end{cases}$$

In this case it is not essential to define the function at the endpoints a_i.

As can be seen easily,

$$\int_{-\infty}^{\infty} f_n(x)\,dx = \sum_{i=1}^{r} \frac{g_i}{a_i - a_{i-1}}(a_i - a_{i-1}) = \sum_{i=1}^{r} g_i = 1.$$

If n is sufficiently large, then the probability of an event $(\alpha < X < \beta)$ lies near the value $\int_{\alpha}^{\beta} f_n(x)dx$, which is the sum of the areas of the rectangles above the interval (α, β).

3.2.3.2 The Empirical Distribution Function. Let X be a continuous random variable with distribution function $F(x)$. Having a sample on X of size n with elements $X_1, X_2, \ldots, X_n$ we define its empirical distribution function $F_n(x)$ in the following way: $F_n(x)$ is the relative frequency of the event $\{X \leqslant x\}$ in the sequence of observations given by the sample. In other words, $F_n(x)$ is the number of sample elements not larger than x divided by n. This means that between any two consecutive sample elements, $F_n(x)$ is of constant value. It can be expressed with the aid of the ordered sample elements (see Fig. 17):

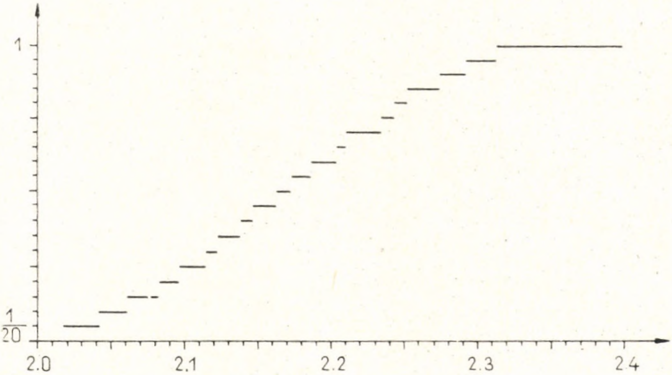


FIG. 17. Empirical distribution function belonging to a sample of size 20

$$F_n(x) = \begin{cases} 0 & \text{if } x < X_1^*, \\ \dfrac{k}{n} & \text{if } X_k^* \leqslant x < X_{k+1}^*, \quad k = 1,2,\ldots, n-1, \\ 1 & \text{if } X_n^* \leqslant x. \end{cases}$$

According to this, $F_n(x)$ is a step function with jumps at the sample elements, continuous on the right. $F_n(x)$ is a random variable for every value of x, its value depending on the (random) sample.

From the definition of the empirical distribution function it follows that, for every fixed x, the random variable $F_n(x)$ tends stochastically to

$F(x)$, as n tends to infinity, i.e., it approaches the value of the theoretical distribution function. This statement is contained in the law of large numbers, $F_n(x)$ being the relative frequency of an event that has the probability $F(x)$. But even a much stronger statement holds, namely, the theorem of Glivenko, called the fundamental theorem of mathematical statistics: Taking the absolute difference $|F_n(x) - F(x)|$ between the theoretical and empirical distribution functions, there is a maximum (or supremum) for each fixed sample; but regarding the sample as an n tuple of random variables, this maximal value is a random variable as well. Let us denote it by Δ_n, i.e.,

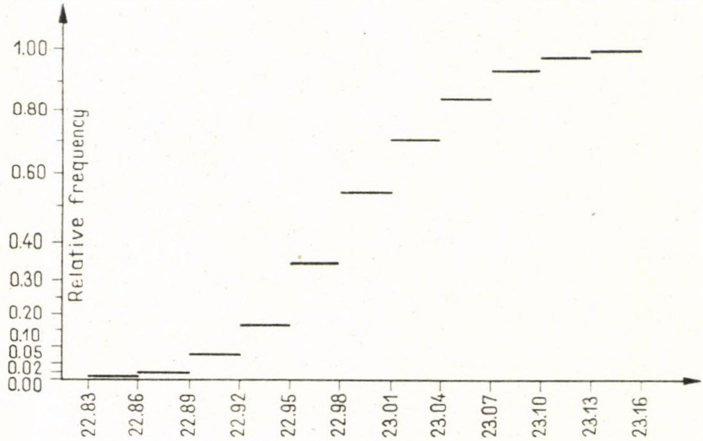


FIG. 18. Empirical distribution function constructed from the grouped data given in Table 6

$$\Delta_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|.$$

Then according to the theorem of Glivenko, this random variable tends to zero with probability 1 as $n \to \infty$:

$$P(\lim_{n\to\infty} \Delta_n = 0) = 1.$$

The great importance of this theorem is due to its consequence: for large enough n the empirical distribution function can be considered a good approximation to the theoretical distribution function.

For a large sample the empirical distribution function can be constructed by grouping the data in the following way (see Fig. 18): In our earlier notations

$$F_n^{(r)}(x)=\begin{cases} 0 & \text{if} \quad x<\dfrac{a_0+a_1}{2}, \\ \displaystyle\sum_{i=1}^{k} g_i & \text{if} \quad \dfrac{a_{k-1}+a_k}{2} \leqslant x < \dfrac{a_k+a_{k+1}}{2}, \quad k=1,2\ldots r-1, \\ 1 & \text{if} \quad \dfrac{a_{r-1}+a_r}{2} \leqslant x. \end{cases}$$

This means that the empirical distribution function based on the grouped data has jumps in the midpoints of the cells, and the size of each jump is equal to the relative frequency g_i of the respective cell. We see that, contrary to the case of $F_n(x)$ where the jump is $1/n$, in the case of $F_n^{(r)}(x)$ it may take on different values, namely multiples of $1/n$; this is sometimes more convenient.

According to the preceding statements, the values $F_n(\beta) - F_n(\alpha)$ or $F_n^{(r)}(\beta) - F_n^{(r)}(\alpha)$ $(\alpha < \beta)$, if n is large enough, can be considered an approximation to or estimate of $F(\beta) - F(\alpha)$, i.e., of the probability of the event $\{\alpha \leqslant x < \beta\}$.

3.3 FUNDAMENTALS OF THE THEORY OF ORDER STATISTICS

3.3.1 Distribution of the Order Statistic

We have already introduced the random variable X_k^*, which is the kth element in order of magnitude $(k = 1, 2, \ldots, r)$, from a sample taken on a *continuous* random variable. Having the independent sample elements $X_1, X_2, \ldots, X_n$, the first order statistic is the smallest among them. Which of the X_i's turns out to be the smallest depends on chance. Owing to the independence and to the identical distribution of all of the sample elements, all of them have the same probability, $1/n$, of becoming X_1^*: $P(X_1^* = X_i) = 1/n$, $i = 1, 2, \ldots, n$. The same statement is true for any X_k^*. This means that knowing the ordered sample $X_1^* < X_2^* < \ldots < X_n^*$ and having lost the original, independent, sample, each of the $n!$ permutations of the ordered sample elements has the same probability $1/n!$ of being the original sample.

We turn now to the distribution function of X_k^*, knowing the common distribution function $F(x)$ of the sample elements. Let $F_{n,k}(x) =$

$P(X_k^* \leqslant x)$ be the distribution function of the kth order statistics; it can be shown that the following relation holds:

$$F_{n,k}(x) = n \binom{n-1}{k-1} \int_{-\infty}^{x} [F(t)]^{k-1} [1 - F(t)]^{n-k} f(t)\, dt$$

$$= \sum_{i=k}^{n} \binom{n}{i} [F(x)]^{i} [1 - F(x)]^{n-i} \quad k = 1, 2, \ldots, n.$$

If, for example, $k = 1$, then

$$P(X_n^* \leqslant x) = F_{n,1}(x) = n \int_{-\infty}^{x} [1 - F(t)]^{n-1} f(t)\, dt = 1 - [1 - F(x)]^{n}$$

and if $k = n$, then

$$P(X_n^* \leqslant x) = F_{n,n}(x) = n \int_{-\infty}^{x} [F(t)]^{n-1} f(t)\, dt = [F(x)]^{n}.$$

A simple representation of $F_{n,k}(x)$ is easy only in certain cases; e.g., if the sample elements are distributed according to the uniform or to the exponential distribution. For certain kinds of distributions (e.g., for the Gaussian distribution) many characteristics of the order statistics are tabulated (see, for example, Sarhan—Greenberg [A 36]).

With respect to the proof of the two forms of $F_{n,k}(x)$ given above we mention the following: both forms can be verified independently on the bases of different probabilistic argumentation; in this way we get a probabilistic proof of the equality of the two forms. This relation will be used later. (The identity of the two forms can be proved by analytic means as well.)

In the second form the term with index i is the probability of the event that among the original sample elements there are exactly i smaller than x and $n - i$ larger than x; since these events are disjoint for different i's, the sum of their probabilities taken from k to n gives exactly the probability of the event $\{X_k^* < x\}$. A similar simple proof can be given for the first form, which contains the density function $f_{n,k}(x)$ of X_k^*:

$$F'_{n,k}(x) = f_{n,k}(x) = n \binom{n-1}{k-1} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x).$$

In Figs 19 and 20 the distribution functions and density functions of the ordered sample elements coming from a Gaussian distribution are shown for the case $n = 5$; the common distribution and density functions of the independent sample elements are given in dotted lines.

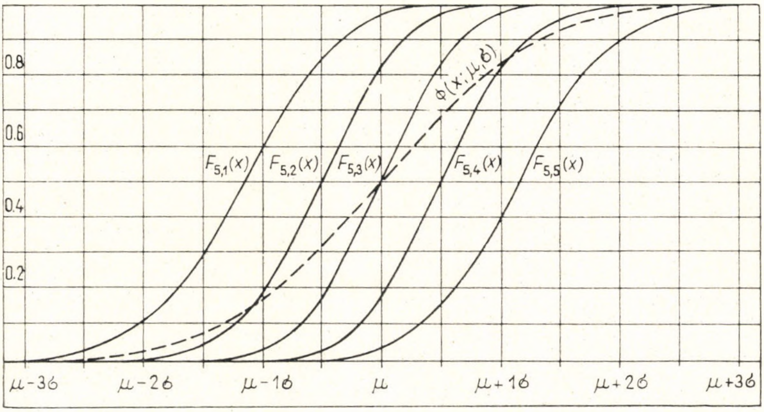


FIG. 19. Distribution functions of the ordered elements of a sample of size 5 taken from a normal population $N(\mu, \sigma)$

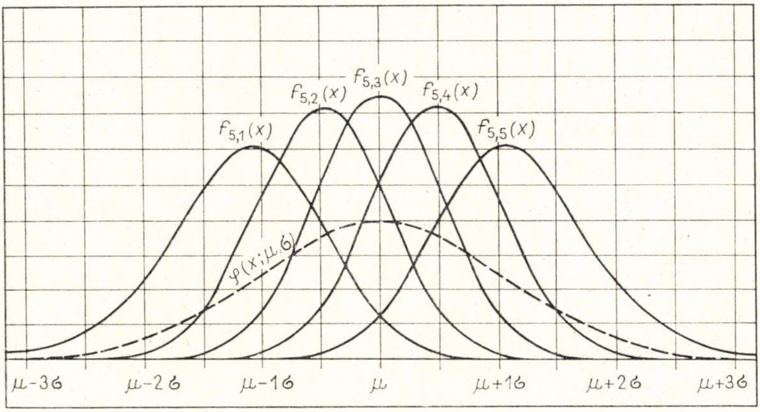


FIG. 20. Density functions of the ordered elements of a sample of size 5 taken from a normal population $N(\mu, \sigma)$

3.3.2 Distribution of the Range

As mentioned earlier, the sample range plays an important role in statistical quality control since it can be determined very easily from the sample. In the case of Gaussian distributions a very simple relation holds between the range and the variance. Knowing the distribution of the range, we can determine probability limits from them. Thus we have a simple quantity for determining whether the variance has changed its value in the course of time.

Let $F(x)$ and $f(x)$ be the common distribution function and the density function, respectively, of the sample elements; then the range $R = X_n^* - X_1^*$ has the distribution function

$$P(R \leqslant y) = K_n(y) = n \int_{-\infty}^{\infty} [F(y+u) - F(u)]^{n-1} f(u)\, du, \quad y \geqslant 0$$

and the density function

$$K_n'(y) = k_n(y) = n(n-1) \int_{-\infty}^{\infty} [F(y+u) - F(u)]^{n-2} f(u)\, f(y+u)\, du$$

$$y \geqslant 0.$$

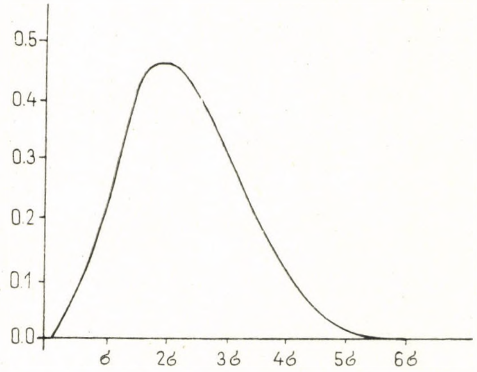


FIG. 21. Density function of the range of a sample of size 5 taken from normal population $N(\mu, \sigma)$

The proof is as simple as that of the distribution functions of the ordered sample elements. Figure 21 shows the density function of the range coming from a sample of size 5 taken from a Gaussian distribution.

We cannot, in general, give an explicit representation of $K_n(y)$ with the aid of elementary or simple functions. Therefore, $K_n(y)$ can be given only in the form of tables for any $F(x)$.

For $F(x) = \Phi(x)$, i.e., for the standard normal law, $K_n(y)$ is tabulated (see, for example, Hald [C 5]). From this table, for example, we obtain for a sample of size $n = 10$ with elements distributed according to $N(\mu, \sigma)$

$$P(R < 4.79\,\sigma) = 0.975.$$

In the case of a Gaussian distribution the expected value and variance of R depends on the sample size n and on the variance σ of the sample elements only:

$$E(R) = d_n\sigma, \quad D(R) = c_n\sigma,$$

where the coefficients d_n and c_n depend on the sample size n only. These are tabulated, and we will give their values in the course of describing certain practical methods.

According to the first relation, the random variable R/d_n fluctuates around the standard deviation σ; this makes it possible to estimate σ by the arithmetic mean of the sample ranges. Let $R_1, R_2, \ldots, R_k$ be ranges computed in independent samples, each of them having size n; then, in the sense presented precisely in the chapter on estimation theory, we may write

$$\sigma \approx s_1 = \frac{1}{d_n} \frac{R_1 + R_2 + \ldots + R_k}{k} = \frac{1}{d_n} \bar{R}.$$

For the variance of the random variable s_1, we have

$$D^2(s_1) = \frac{1}{d_n^2} \frac{\sum_{i=1}^{k} D^2(R_i)}{k^2} = \frac{1}{k} \frac{e_n^2}{d_n^2} \sigma^2.$$

This means that if k is sufficiently large, then the variance, and consequently the fluctuation of s_1, is small, and the estimation for σ given above will be suitable.

3.3.3 Theorems of Kolmogorov and Smirnov

Let us again denote by $F_n(x)$ the empirical distribution function of a sample of size n and by $F(x)$ the common distribution function of the sample elements. As mentioned in Section 3.2.3.2, the maximum distance between the empirical and the theoretical distribution functions, i.e., the random variable

$$\Delta_n = \sup_{-\infty < x < \infty} | F_n(x) - F(x) |,$$

tends to 0 with probability 1; this was the statement of Glivenko's theorem.

Kolmogorov and Smirnov considered the question of how fast this convergence is. They found that Δ_n tends to zero at the same rate as $1/\sqrt{n}$ does. Before giving a precise formulation of their statement we introduce the *one-sided* deviation of the empirical and the theoretical distribution functions:

$$\Delta_n^+ = \sup_{-\infty < x < \infty} [F_n(x) - F(x)].$$

Now the following limiting distribution theorems hold (due to Smirnov and Kolmogorov, respectively):

$$\lim_{n\to\infty} P\{\sqrt{n} \sup_{-\infty<x<\infty} [F_n(x) - F(x)] \leqslant y\} = 1 - e^{-2y^2}, \quad y \geqslant 0$$

and

$$\lim_{n\to\infty} P[\sqrt{n} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \leqslant y] = \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2y^2}, \qquad y > 0.$$

The exact distribution laws of the random variables Δ_n and Δ_n^+ are rather complicated. The distribution of Δ_n^+ has been determined independently by Smirnov [B 73] and by Birnbaum and Tingey [B 11]. The latter authors have also tabulated the limiting distribution function.

In order to show how these limiting distribution laws work, let y_ε be the value for which the distribution function of $\sqrt{n}\Delta_n$ is close to $1 - \varepsilon$, n being large; i.e.,

$$\sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2y_\varepsilon^2} = 1 - \varepsilon \quad (\varepsilon \text{ small}).$$

In this case, with probability near to $1-\varepsilon$, which is close to 1, the following relation holds:

$$\sqrt{n} \mid F_n(x) - F(x) \mid < y_\varepsilon.$$

This relation can be interpreted in two different ways, according to whether the theoretical distribution function is known or not. In the first case we write the relation in the

$$F(x) - \frac{y_\varepsilon}{\sqrt{n}} < F_n(x) < F(x) + \frac{y_\varepsilon}{\sqrt{n}},$$

which gives a strip around the theoretical distribution function of width $2y_\varepsilon/\sqrt{n}$, containing the empirical distribution function with probability close to $1 - \varepsilon$. Therefore we are able to decide of whether or not the sample elements really come from a population with distribution function $F(x)$. On the other hand, if $F(x)$ is unknown, then we have a strip with the same width $2y_\varepsilon/\sqrt{n}$ around $F_n(x)$, in which $F(x)$ must lie "with probability near to 1":

$$P\left(F_n(x) - \frac{y_\varepsilon}{\sqrt{n}} < F(x) < F_n(x) + \frac{y_\varepsilon}{\sqrt{n}}\right) \approx 1 - \varepsilon.$$

3.3.4 Distributions of the Maximum Distances Between Two Empirical Distribution Functions

In statistical control work the comparison of the quality or of some characteristics of two products is of importance, and the need for such comparison occurs very frequently. For this reason we will state the following distribution laws in which two samples play a role. The application of these theorems will be discussed in the section on testing statistical hypotheses.

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent samples, i.e., observations on the random variables X and Y, respectively. For the distribution functions we use the notations

$$P(X \leqslant x) = F(x), \; P(Y \leqslant x) = G(x).$$

We will assume throughout this section that $G(x) \equiv F(x)$, which means that the $m + n$ random variables of both samples are identically distributed. Let us construct the empirical distribution functions $F_n(x)$ and $G_m(x)$ belonging to the first and second samples. According to Glivenko's theorem, both empirical distribution functions converge with probability 1 to the common distribution function $F(x)$. We will consider now the maximal deviations of the two empirical distribution functions; the one-sided deviation is

$$D_{n,m}^{+} = \sup_{-\infty<x<\infty} [F_n(x) - G_m(x)],$$

while the absolute (two-sided) deviation is

$$D_{n,m} = \sup_{-\infty<x<\infty} |F_n(x) - G_m(x)|.$$

It is not difficult to prove that Glivenko's theorem holds for these random variables as well, i.e., that the following relation is valid:

$$P(\lim_{\substack{n\to\infty \\ m\to\infty}} D_{n,m}^{+} = 0) = P(\lim_{\substack{n\to\infty \\ m\to\infty}} D_{n,m} = 0) = 1 .$$

If at the same time $n \to \infty$, $m \to \infty$, $n/m \to c \neq 0$, then the following limiting distribution theorems, due to Kolmogorov and Smirnov, hold:

$$\lim_{\substack{n\to\infty \\ m\to\infty}} P\left[\left(\frac{nm}{n+m}\right)^{1/2} \sup_{-\infty<x<\infty} [F_n(x) - G_m(x)] < y\right] = 1 - e^{-2y^2}, \quad y \geqslant 0 .$$

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\left[\left(\frac{nm}{n+m}\right)^{1/2} \sup_{-\infty<x<\infty} F_n(x)-G_m(x)<y\right]=$$

$$\sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2y^2}, \quad y>0.$$

By means of these distribution laws we can determine a value y_ε^+ (y_ε) so that if both samples indeed come from the same distribution, then the maximal distance of the two empirical distribution functions does not exceed y_ε^+ (y_ε) with probability close to $1-\varepsilon$. This holds, however, for large values of n and m only. For moderate or even small samples, in the case of equal sample sizes $(m=n)$ the following distribution law, due to Gnedenko and Korolyuk, is valid. (It can be seen both $F_n(x)$ and $G_n(x)$ are fractions with the common denominator n, and the same is true for their difference.)

$$P\left(\max_{-\infty<x<\infty}\left[F_n(x)-G_n(x)\right]\leqslant\frac{k}{n}\right)=1-\frac{\binom{2n}{n-k-1}}{\binom{2n}{n}},$$

$$k=0,1,2,\dots,n.$$

$$P\left(\max_{-\infty<x<\infty}\left|F_n(x)-G_n(x)\right|\leqslant\frac{k}{n}\right)=\sum_{i=-\infty}^{\infty}(-1)^i\frac{\binom{2n}{n-i(k+1)}}{\binom{2n}{n}},$$

$$k=1,2,\dots,n.$$

In the former case $D_{n,m}^+$ can take on the value 0 with positive probability, while in the latter case this can happen with probability 0 only.

4 FUNDAMENTALS OF MATHEMATICAL STATISTICS

4.1 BASIC PROBLEMS OF MATHEMATICAL STATISTICS

4.1.1 The Two Main Branches of Mathematical Statistics

Mathematical statistics can be divided into two main branches: ***statistical estimation theory*** and the ***testing of statistical hypotheses***. Let us consider each in turn.

In industrial quality control we find a large number of random quantities. In most cases we do not know the distributions and theoretical characteristics; our aim is to determine some of these quantities. This is usually accomplished with the aid of observations on the random variables; we obtain information from the samples. According to the laws of large numbers, an "exact" determination of a probability, an expected value, etc. would require an "infinite" number of observations. Having samples of finite size, we can do no more than *estimate* the theoretical values in question. The sample characteristics, or, in other words, the various statistics introduced in the previous section, serve the purposes of statistical estimation. In Section 4.2 we will make clear what an estimation is, how to choose among various *estimators*, and how to work with estimators.

For a good estimation of theoretical quantities, a fairly large sample is sometimes needed. The preliminary sampling used in statistical quality control to determine the distribution of the random quantity in question is therefore based on large samples. On the other hand, quality control is carried out in the course of production with the aid of small samples, mainly for practical reasons. In the case of quality control, however, we have already determined the distribution or the characteristic. We make the assumption that this distribution has not changed during production; our task is to verify this assumption. Knowledge of the original distribution gives us information on the customary behavior of the sample characteristics.

Thus, the testing of statistical hypotheses, based on samples that aid us in checking the validity of certain assumptions, is the second main branch of mathematical statistics; it will be treated in Section 4.3.

As mentioned earlier, the final result of a statistical inference is always a *decision*: we decide to accept for the value of a theoretical quantity the observed value of its estimator, without demanding that it be exactly the true value; or we decide to accept or to reject an assumption concerning a certain distribution, without saying that our statement is true beyond doubt. By using statistical methods we are able to control the possible errors of our decisions—their extent and the frequency with which they occur. Our aim is to devise methods having the smallest possible errors or to allow only errors that are still economical.

4.1.2 Parametric and Nonparametric Problems in Mathematical Statistics

Let the random variable X be the length of a cylinder produced by a machine. Suppose that we know that X is distributed according to the Gaussian law:

$$P(X \leqslant x) = \Phi(x; \mu, \sigma_0) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma_0^2}\right) dt,$$
$$-\infty < x < \infty.$$

The fact that the fluctuation of the values is subject to the normal law is a property of the machine, while the variance $D^2(x) = \sigma_0^2$ is very often determined by the operation of the machine and is known from long experience. For the length of the cylinder we have a prescribed or nominal value μ. At the beginning of production we try to set the machine so that the cylinders produced will have lengths that fall within the tolerance limits, which are centered around μ_0. This means that the random fluctuation should take place around μ_0. Whether this is actually the case can be determined by estimating the expected value μ of the distribution.

Expressed in the language of mathematical statistics:

We do not know the distribution function of the random variable X, but we know that this distribution function is of a definite type or family of distribution functions. This means that for the unknown distribution function a certain set of distribution functions comes under consideration, the members of which are determined by one (or by a finite number of) *parameters*.

If, in the above case, neither the expected value nor the variance are known, then the set of distributions in question is the family of the Gaussian distribution functions:

$$\{\Phi(x; \mu, \sigma):\ 0 < \mu < \infty,\ 0 < \sigma\}.$$

Here, according to the meaning of the quantity in question, only positive expected values come into consideration.

When the value σ_0^2 of the variance is known, the set of possible distributions reduces to

$$\{\Phi(x; \mu, \sigma_0):\ 0 < \mu < \infty\}.$$

A second example: From experience we know that the yarn breakages ν during a shift are distributed according to the Poisson law. Having a new kind of yarn we still know the type of distribution, but its parameter is unknown:

$$\{F(k, \lambda);\ 0 < \lambda < \infty\},$$

where

$$F(k, \lambda) = P(\nu \leqslant k) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots\ .$$

We speak, in general, of *parametric problems* if the underlying distribution belongs to a set of distribution functions, the members of which are uniquely determined by a finite number of parameters. In such cases the preliminary sampling concerns the estimation of the unknown parameters; then, at a later time (or at fixed intervals), we test the hypothesis that the values of the parameters have not changed in the meantime. The former is carried out by large samples; the latter is based on small samples.

There are, however, sets of distribution functions whose members cannot be determined by a finite number of parameters (e.g., the set of all continuous distribution functions or the set of all continuous distribution functions having a given median, etc.). This is called the *nonparametric case,* and the statistical methods used for this kind of problem are the *nonparametric methods.* It frequently occurs that in the absence of knowledge of the distribution type we use nonparametric methods.

If we know only that the distribution function is continuous, the set of distribution functions that comes into consideration is the set of all continuous distribution functions:

$$\{F(x)\colon\ F(x)\ \text{continuous}\}.$$

In the case of continuous distribution functions, nonparametric methods based on *order statistics* are used. We wish to remark again that these methods are employed even in parametric problems when they are more economical than the corresponding parametric methods.

4.2 FUNDAMENTALS OF THE THEORY OF ESTIMATION

4.2.1 The Estimation Problem

In many practical situations the following two types of estimation problems arise.

A certain technical constant, say a, which is, from the statistical point of view, a theoretical quantity, has to be determined by means of measurement. Such constants may be, for example, the length or the diameter of a given product, the electrical resistance of a given device, the concentration of a given solution, etc. The result X of the measuring procedure is a random variable whose distribution depends on a (and perhaps on further constants, which will be left out of consideration here):

$$P(X \leqslant x) = F(x;\ a),\ a \in S,$$

where S denotes the set of values the unknown parameter a may take on. In the language of mathematical statistics we have to estimate the parameter a out of a sample $X_1, X_2, \ldots, X_n$ taken on X.

In the other case the quantity in question is a random variable itself: for example, the diameter of cylinders coming from a given machine, the temperature in a boiler during a long period, etc. In such cases we are interested in the (theoretical) average value, and/or the dispersion of X, etc. This means that we have to estimate the expected value $E(X)$ and/or $D^2(X)$, and perhaps other (constant) quantities that can be expressed with the aid of the distribution function of X.

In estimating a constant value there are various aspects to consider. If we wish to have an estimator whose value can be used instead of the unknown parameter in formulas, then the estimator should have one given value. In this case we speak of *point estimation.*

But knowing that our estimator is subject to error, sometimes we would like to have some information on the deviation from the true value. In this case we have to construct an interval that contains the unknown parameter, at least with high probability. This method is called *interval estimation.*

With respect to the reliability of both types of estimation procedures the procedures should satisfy certain requirements. The estimation methods and their properties will be given in the next section.

4.2.2 Point Estimation

We wish to introduce the main concepts of estimation theory with the aid of the example treated earlier; following this, the general formulations will be given.

Let us suppose that the random variable X is distributed according to the normal law $\Phi(x; \mu, \sigma_0)$ with known variance $D^2(X) = \sigma_0^2$ and the unknown expected value μ, which is therefore the parameter to be estimated. To estimate μ we consider a sample of size n on X with elements $X_1, X_2, \ldots, X_n$. Then we have for $i = 1, 2, \ldots, n$

$$P(X_i \leqslant x) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma_0^2}\right) dt, \quad -\infty < x < \infty .$$

According to our considerations in Section 3.2.2.1 concerning the sample mean, it is convenient to estimate μ with the aid of the arithmetic mean. Denoting by the sign $\sim$ the statistical estimation of a (constant) parameter with the aid of a statistic, the estimation can be written in the form

$$\mu \sim \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

It is obvious that a good estimator should have the properties considered in the next sections.

4.2.2.1 Unbiasedness. In principle, each statistical method is to be considered as a method used repeatedly. In single cases the estimator will differ from the estimated value, but in a large number of cases the deviations must not prefer either direction. This requirement will be fulfilled if the expected value of the estimator equals the parameter estimated; in our example this indeed holds:

$$E(\overline{X}) = \mu \text{ for each value of } \mu. \tag{11}$$

This relation must be true regardless of the actual value of μ; if, for example, we produce cylinders of length $\mu' > \mu$, then X will fluctuate around μ', and the expected value will be equal to μ'.

In general: Let us have a random variable X whose distribution function depends on certain parameters $a_1, a_2, \ldots, a_k$:

$$P(X < x) = F(x;\ a_1, a_2, \ldots, a_k).$$

The sample statistic $\alpha = \alpha(X_1, X_2, \ldots, X_n)$ is called an *unbiased estimator* for the parameter a if

$$E(\alpha) = a$$

for each possible set of the values $a_1, a_2, \ldots, a_k$. The parameter a may be one of the a_i's, but it may also be a function of them; e.g., $a = a_2 - a_1^2$.

The validity of Eq. (11) follows from the fact that μ is the common expected value of the X_i's (see Section 2.6).

Consequently, it holds in general that the sample mean is an unbiased estimator for the expected value of the random variable if the first moment of the distribution exists.

The same is true for the variance of a random variable: *If the second moment of the random variable X exists, then the sample variance is an unbiased estimator for the (theoretical) variance:*

$$D^2(X) \sim s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}, \quad E(s^2) = D^2(X), \tag{12}$$

regardless of any further parameters that may occur in the distribution of X. For unbiasedness we divide by $(n - 1)$ instead of by n in the expression for s^2.

On the other hand, *there does not exist an unbiased estimator for the (theoretical) standard deviation* $D(X)$ in general. For certain families of distributions, however, an unbiased estimator for $D(X)$ exists.

Let the random variable X be *normally distributed with parameters* μ *and* σ^2:

$$P(X \leqslant x) = \Phi(x;\ \mu, \sigma^2).$$

Then the statistic

$$s_1 = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \sqrt{n-1}\, \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\, s$$

is an unbiased estimator for the standard deviation $D(X)$:

$$\sigma \sim s_1 \qquad E(s_1) = \sigma. \tag{13}$$

We omit the proof of relations (12) and (13). The first relation can be obtained by direct calculation. For the second, the distribution of s_1 or s is needed in the case of the normal distribution; in this case $(n - 1)s^2/\sigma^2$ is χ^2-distributed with parameter $(n - 1)$, as is given in Section 2.7.6; from this the distribution and expected value of s_1 can be derived.

4.2.2.2 Efficiency. Once we have one unbiased estimator for a parameter occurring in the distribution function of a random variable, in most cases we have many such estimators. Let X again be distributed according to $N(\mu, \sigma_0)$ with σ_0 known. Let $c_1, c_2, \ldots, c_n$ be real known constants with $\sum_{i=1}^{n} c_i = 1$. As can be shown very easily, the random variable

$$\bar{X}_c = \sum_{i=1}^{n} c_i X_i$$

also has the expected value μ; in this way we have a set of unbiased estimators for μ according to the possible choices of the c_i's. $\bar{X}$ itself belongs to this set with $c_1 = c_2 = \ldots = c_n = 1/n$. At this point the question arises: How should we choose the most suitable estimator out of this set of estimators or perhaps out of an even larger set of unbiased estimators?

Because estimators are random variables, we prefer those that are subject to smaller fluctuations around the estimated quantity; as a measure of fluctuation we accept the variance. The *unbiased estimator* $\alpha_1 = \alpha_1(X_1, X_2, \ldots, X_n)$ *will be called (uniformly) more efficient for the parameter a than the unbiased estimator* $\alpha_2 = \alpha_2(X_1, X_2, \ldots, X_n)$ if the variance of α_1 is less than the variance of α_2:

$$D^2(\alpha_1) \leqslant D^2(\alpha_2)$$

for all possible values of the parameter that occur in the set of distributions in question.

Let the set of distributions in question have first and second moments; then among *all linear estimators for the expected value, the sample mean has the smallest variance*, i.e., it is the most efficient estimator for the expected value in this set of estimators:

$$D^2(\bar{X}) \leqslant D^2(\sum_{i=1}^{n} c_i X_i), \quad \sum_{i=1}^{n} c_i = 1.$$

This follows from the following relations: on the one hand,

$$D^2(\bar{X}) = \frac{D^2(X)}{n},$$

and on the other hand, the X_i's being independent,

$$D^2(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} c_i^2 \, D^2(X_i) = D^2(X) \sum_{i=1}^{n} c_i^2.$$

But, according to the inequality of Schwarz, if $\sum_{i=1}^{n} c_i = 1$, then $\sum_{i=1}^{n} c_i^2 \geqslant 1/n$ and equality holds if and only if $c_1 = c_2 = \ldots = c_n = 1/n$.

In the case of certain families of distributions, even more than that can be said: *For families of normal, exponential, or Poisson distributions the sample mean has the smallest variance among all possible unbiased estimators for the expected value.*

To show that this statement is not true in general, we mention the case of uniform distributions. Let X be distributed according to the uniform distribution (see Section 2.9.11) in the interval (a, b). The expected value of X is

$$E(X) = \frac{a+b}{2}.$$

Let the length $d = b - a$ of the interval be known but the location of it, i.e., the expected value, unknown. According to our earlier theorem the sample mean $\bar{X}$ with variance $D^2(\bar{X}) = D^2(X)/n = (b-a)^2/12\,n$ is a more efficient estimator than any $\overline{X}_c$. But taking into consideration the ordered sample we can show that

$$E\left(\frac{X_1^* + X_n^*}{2}\right) = \frac{a+b}{2}, \quad D^2\left(\frac{X_1^* + X_n^*}{2}\right) = \frac{(b-a)^2}{2(n+1)(n+2)}.$$

This means that the arithmetic mean of the first and last order statistic has a variance that is of the order of magnitude $1/n^2$, while that of the sample mean, is $1/n$.

An unbiased estimator whose variance is minimal among all unbiased estimators is called the *minimum variance unbiased estimator* or the ***best unbiased estimator***. The question of the existence of such estimators will be treated in Section 4.3.2.

4.2.2.3 Consistency. Returning to our previous example, let X be distributed according to $N(\mu, \sigma_0^2)$ with σ_0 known. For estimating $E(X) = \mu$ we have chosen the statistic $\overline{X}$, which is, as mentioned, the best unbiased estimator. But X has, as an estimator, even a further favorable property: It tends stochastically to μ as n tends to infinity:

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \Rightarrow \mu, \quad n \to \infty.$$

We recall the definition of stochastic convergence: The sequence $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ of random variables tends stochastically to the value a if for any given $\varepsilon > 0$ the probability of the event $\{|\alpha_n - a| > \varepsilon\}$ tends to zero as n tends to infinity. In the above case this means that for any $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) = 0. \tag{14}$$

This limiting relation is true not only in the Gaussian case but in the case of all random variables having first moments; e.g., the exponential, Poisson, Gamma, uniform law, etc. But a relation such as Eq. (14) does not hold if the random variable has a Cauchy distribution (Student with $n = 1$; see Section 2.9.8), since in this case no first moment exists.

An unbiased estimator α_n for the parameter a is called *strongly consistent* if its variance tends to zero as n tends to infinity:

$$E(\alpha_n) = a, \; D^2(\alpha_n) \to 0, \text{ if } n \to 0.$$

In the case of random variables having a second moment (e.g., the Gaussian, exponential, Poisson laws) the sample mean is a strongly consistent estimator for the expected value.

4.2.2.4 Sufficiency. Sufficiency is one of the central concepts of mathematical statistics.

Again, let the lengths X of the cylinders be distributed according to $N(\mu, \sigma_0)$ with σ_0 known. Considering the sample $X_1, X_2, \ldots, X_n$ on X, let us assume that we have carried out an observation on this sample that has resulted in $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. The estimation of μ will be carried out by means of the sample mean:

$$\mu \sim \bar{x} = \frac{x_1 - x_2 + \ldots + x_n}{n} .$$

Suppose now that we have lost the list of values of the single sample elements but we have not lost the value of $\bar{x}$; in this case we still have an estimator for μ. However, we may raise the question of whether the knowledge of all sample elements contained more information regarding μ than does $\bar{x}$. This is a reasonable question, since to any fixed $\bar{x}$ there exists an infinite set of n-tuples $y_1, y_2, \ldots, y_n$ with $(y_1 + y_2 + \cdots + y_n)/n = \bar{x}$, and in using $\bar{x}$ we do not utilize the fact that the result of our observation was exactly $(x_1, x_2, \ldots, x_n)$. To examine this question we have to construct the joint distribution function of the sample elements under the condition that their arithmetic mean is the given value $\bar{X} = \bar{x}$:

$$P(X_1 \leqslant y_1, X_2 \leqslant y_2, \ldots, X_u \leqslant y_n \mid \bar{X} = \bar{x}).$$

It may well be imagined that by estimating μ from this distribution we could get a better estimator or in a certain sense further information on μ. But it turns out that *the conditional distribution of the sample elements under the condition that the sample mean is given does not depend on the expected value* μ; this means that it does not contain any further information about μ. This is true for random variables distributed according to the normal law, and it holds for the exponential and Poisson case as well. Knowing the observed value of the statistic, we cannot draw any further information on the unknown parameter from the single sample elements.

The conditional joint density function of the sample elements has the form

$$f(x_1 x_2, \ldots x_n \mid \bar{X} = \bar{x}) = \begin{cases} \dfrac{f(x_1, x_2, \ldots, x_n; \mu, \sigma_0)}{f(\bar{x}\,; \mu, \sigma_0)} & \text{if } \sum\limits_{i=1}^{n} x_i = n\bar{x} , \\ 0 & \text{if } \sum\limits_{i=1}^{n} x_i \neq n\bar{x} , \end{cases} \tag{15}$$

where $f(x_1, x_2, \ldots, x_n;\ \mu, \sigma_0)$ is the joint density function of the sample elements and $f_n(\bar{x};\ \mu, \sigma_0)$ is the density function of the sample mean. Since in the normal case

$$f(x_1, x_2, \ldots, x_n; \mu, \sigma_0) = \prod_{i=1}^{n} \varphi(x_i; \mu, \sigma_0)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma_0}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{2\sigma_0^2}\right)$$

and

$$f_n(\bar{x}; \mu, \sigma_0) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{n(\bar{x} - \mu)}{2\sigma_0^2}\right),$$

our above statement can be obtained very easily. On the other hand, from relation (15) it follows that the joint density function of the sample elements has the form

$$f(x_1, x_2, \ldots, x_n;\ \mu, \sigma_0) = f(x_1, x_2, \ldots, x_n \mid \bar{X} = \bar{x})\ f_n(\bar{x};\ \mu, \sigma_0).$$

This is the so-called Neyman factorization property in the presence of sufficient statistics. We shall use this as the definition of sufficient statistics. What we have to note here is that the first factor does not depend on the parameter μ, while the second factor depends on the sample elements through the (sufficient) statistic $\bar{X}$ only.

Let X now be a random variable whose distribution is an element of a family of distributions depending on a real parameter a:

$$P(X \leqslant x) = F(x;\ a) = \int_{-\infty}^{x} f(t, a)dt, \quad a \in S,$$

where S denotes the set of points on the real axis that are the possible values of a (e.g., the whole real line, a finite or infinite interval, the integer points, etc.).

The statistic $\alpha_n = \alpha_n(X_1, X_2, \ldots, X_n)$ is called sufficient with respect to the parameter a or with respect to the family $\{F(x;\ a),\ a \in S\}$ of distributions if, for the joint density of sample elements, the factorization property holds:

$$f(x_1, x_2, \ldots, x_n;\ a) = \prod_{i=1}^{n} f(x_i, a) = f_1(x_1, x_2, \ldots, x_n)\ f_2(y;\ a),$$

where f_1 depends on the variables $x_1, x_2, \ldots, x_n$, but not on the parameter a, while $f_2(y; a)$ depends on the X_i's through $y = \alpha_n(X_1, X_2, \ldots, X_n)$ only. The functions f_1 and f_2 can be determined in the following way:

f_1 is the conditional density function of the sample elements under the condition that $\alpha_n(X_1, X_2, \ldots, X_n) = y$ in which the parameter a does not occur, f_2 is the density function of the random variable $\alpha_n(X_1, X_2, \ldots, X_n)$.

In the case of the exponential distribution we have

$$P(X \leqslant x) = 1 - e^{-ax}, \quad 0 < a < \infty.$$

The density function of X is

$$f(x, a) = ae^{-ax},$$

and the joint density of the independent sample elements is

$$\prod_{j=1}^{n} f(x_j; a) = a^n \exp\,(-a \sum_{j=1}^{n} x_j).$$

It follows from this that with $f_1(x_1, x_2, \ldots, x_n) = 1$ and $f_2(y; a) = a^n e^{-any}$, where $y = (1/n) \sum_{j=1}^{n} x_j$, the factorization property holds. Consequently, the statistic $\alpha_n = \bar{X}$ is sufficient for the parameter a or for the family $\{F(x;\ a) = 1 - e^{-ax},\ 0 < a < \infty\}$ of distribution functions.

A similar simple derivation leads to the result that in the case of the Poisson distribution the sum $\sum_{i=1}^{n} X_i$ is a sufficient statistic for the parameter of the distribution.

Let us assume that in the Gaussian case the parameters μ and σ both are unknown. Then the following statement holds: The statistic $(\bar{X}, s^2)$, i.e., the vector variable consisting of the sample mean and sample variance, is sufficient for the parameter vector (μ, σ^2), i.e., for the family $\{\Phi(x;\ \mu, \sigma),\ -\infty < \mu < \infty,\ 0 < \sigma < \infty\}$ of Gaussian distributions.

The definition of sufficiency in the case of vector parameters is analogous to the one-parameter case: the conditional joint distribution (density) function of the sample elements, under the condition that the observed values of the two components are given, is independent of the parameters:

$$P(X_1 \leqslant x_1,\ X_2 \leqslant x_2, \ldots, X_n \leqslant x_n \mid \bar{X} = \bar{x}, s^2 = y) = H(x_1, x_2, \ldots, x_n),$$

where H does not depend on μ and σ.

We wish to remark that, knowing the expected value μ of the normally distributed random variable, for σ^2 the statistic $\sum_{i=1}^{n}(X_i - \mu)^2$ is sufficient, instead of $\sum_{i=1}^{n}(X_i - \bar{X})^2$.

In estimating a parameter or testing statistical hypotheses concerning parametric families of distribution functions, we prefer sufficient statistics if possible. These lead to estimators with minimum variance or to the most favorable test functions if such exist.

In the case of random variables having Weibull distribution, i.e.,

$$P(X < x) = 1 - e^{-\alpha x^{\beta}}, \; x \geqslant 0, \; (\alpha > 0, \; \beta > 1),$$

no sufficient statistic exists for the parameter β.

4.2.2.5 The Cramér–Rao Inequality. Let X be a random variable with its distribution function containing a parameter a, which varies within a certain subset S of the real axis:

$$P(X \leqslant x) = F(x;\, a), \quad a \in S.$$

Let us assume now that perhaps no unbiased estimator exists for a, but that we can construct an unbiased estimator for a function $\psi(a)$ of a. This includes the case when a itself has an unbiased estimator ($\psi(a) = a$).

Usually we have not one but several unbiased estimators for a. Our aim is to select an estimator with a variance as small as possible. The question arises: How small can we make the variance? Such a limit exists, depending on the distribution. For certain types of distribution this limit is expressed in the Cramér–Rao inequality, which is as follows:

Let $(X_1, X_2, \ldots, X_n)$ be a sample, i.e., random variables with the common density functions $f(x;\, a)$, $a \in S$, and let $\alpha_n = \alpha_n(X_1, X_2, \ldots, X_n)$ be an unbiased estimator for $\psi(a)$:

$$E(\alpha_n) = \psi(a) \text{ for all } a \in S.$$

We assume that the set of points on the real axis where $f(x;\, a) > 0$ does not depend on a, and furthermore that

$$\frac{d}{da}\int_{-\infty}^{\infty} f(x; a)\, dx = \int_{-\infty}^{\infty} \frac{df(x; a)}{da}\, dx = 0,$$

$$\psi'(a) = \frac{d}{da}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \alpha_n(x_1\, x_2, \ldots, x_n) \prod_{i=1}^{n} f(x_i; a)\, dx_i$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \alpha_n(x_1, x_2, \ldots, x_n) \frac{d}{da} \prod_{i=1}^{n} f(x_i; a)\, dx_i$$

holds. Then for the variance of α_n the following is valid:

$$D^2(\alpha_n) \geqslant \frac{[\psi'(a)]^2}{nI},$$

where

$$I = E\left(-\frac{d^2 \log (X; a)}{da^2}\right).$$

$[\psi'(a)]^2/(nI)$ is the so-called *information limit* for $\psi(a)$. If $\psi(a) = a$, then we have simply $1/(nI)$ for the limit. Estimators for which the equality sign can be attained are *efficient estimators* or *best unbiased estimators.*

The condition that the set of x values for which $f(x;\ a) > 0$ does not depend on a is fulfilled in cases when $f(x;\ a) > 0$ on the whole real axis or on the positive half-line $x > 0$, or on some fixed interval, the endpoints of which do not depend on a. This condition is not fulfilled for example, in the case of the uniform distribution (Section 2.9.1). As we have already seen in connection with the concept of efficiency, the expected value of the uniform distribution can be estimated with a smaller variance than given by the Cramér–Rao inequality; it can be proved that the estimator given there is the best unbiased estimator.

It is in general true that if an unbiased estimator with the smallest variance exists at all, then only one such estimator exists.

The density functions for which equality can be achieved in the Cramér–Rao inequality have the form

$$f(x;\ a) = h(x)\ \exp\,[A(a)t(x) + B(a)],$$

where h and t depend on x, while A and B depend on a only. In this case the factorization property is fulfilled, i.e., sufficient statistics exist.

We will not go into more detail about this interesting field; some consequences for certain types of distributions will be mentioned in the next section.

Finally, we wish to remark that we have formulated here the Cramér–Rao inequality for continuous random variables; it is, however, valid for discrete variables too.

4.2.2.6 Estimation of Some Important Parameters. (a) *Estimation of the probability*. Let A be an event connected with a random experiment with unknown probability p:

$$P(A) = p, \quad 0 \leqslant p \leqslant 1.$$

Let us denote by $\varkappa_n$ the number of occurrences of the event A in n independent trials. As we know, $\varkappa_n$ is distributed according to the binomial law:

$$P(\varkappa_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

(we point out that $\varkappa_n$ is a sufficient statistic). Since $E(\varkappa_n) = np$, we have an unbiased estimator for the probability

$$\frac{\varkappa_n}{n} \sim p, \quad E\left(\frac{\varkappa_n}{n}\right) = p$$

with minimum variance

$$D^2\left(\frac{\varkappa_n}{n}\right) = \frac{p(1-p)}{n}.$$

We can say that *the relative frequency is the uniquely best unbiased estimator for the probability and it is strongly consistent as well.*

(b) *Estimation of the distribution and the density function.* The estimation of the distribution function and of the density function is relatively simple in the parametric case, i.e., when the set of possible distributions is given by means of a finite number of parameters. In this case we obtain estimators for the distribution and density functions by replacing the unknown parameters by their estimators. For example, let X be distributed according to $N(\mu, \sigma)$ with μ and σ unknown, and let us use as estimators $\mu \sim \overline{X}$ and $\sigma \sim s_n$. Then

$$\varphi(x; \mu, \sigma) \sim \varphi(x; \overline{X}, s_n) = \frac{1}{\sqrt{2\pi}\, s_n} \exp\left(-\frac{(x-\overline{X})^2}{2s_n^2}\right)$$

and

$$\Phi(x;\ \mu, \sigma) \sim \Phi(x;\ \overline{X}, s_n) = \int_{\infty}^{x} \varphi(t;\ \overline{X}, s_n)\, dt.$$

These estimators are not unbiased; in certain cases, however, unbiased estimation can be given; in the above Gaussian case, for example, if $\sigma = \sigma_0$ is known, then the estimation

$$\Phi(x; \mu, \sigma_0) \sim \Phi\left[\frac{x-\overline{X}}{\sigma_0}\left(\frac{n}{n-1}\right)^{1/2}\right]$$

is unbiased.

Furthermore, if μ and σ are both unknown, then the estimator

$$\Phi(x; \mu, \sigma) \sim \begin{cases} 0, & \text{if } \tau \leqslant 0 \\[1ex] \dfrac{\Gamma(n-2)}{\left[\Gamma\left(\dfrac{n-2}{2}\right)\right]^2} \displaystyle\int_0^{\tau} [u(1-u)]^{\frac{n-1}{2}}\, du, & \text{if } 0 \leqslant \tau \leqslant 1 \\[1ex] 1, & \text{if } 1 \leqslant \tau \end{cases}$$

is unbiased with

$$\tau = \frac{1}{2} - \frac{\overline{X} - x}{s_n} \frac{\sqrt{n}}{2(n-1)} .$$

This result is due to Kolmogorov [B 45], and the necessary values occurring in both estimations are tabulated, for some selected values of n, in Liebermann–Resnikoff [B 47].

In the nonparametric case, however—i.e., if we do not know anything about the type of distribution—we can obtain the estimators of the distribution function and the density function in the following way. Denoting by $F(x)$ the distribution function of x, $F(x)$ is the probability of the event $\{X \leqslant x\}$; the empirical distribution function $F_n(x)$ constructed from a sample of size n is the relative frequency of the mentioned event; according to the above-mentioned relation between the probability and the relative frequency, the following statement is true:

For every value of x the empirical distribution function is the best unbiased estimator for the theoretical distribution function:

$$F(x) \sim F_n(x), \quad E[F_n(x)] = F(x) \text{ for all } x$$

with variance

$$D^2[F_n(x)] = \frac{F(x)\left[1 - F(x)\right]}{n} \quad \text{for all } x .$$

From the last relation we have that $F_n(x)$ is strongly consistent for all x. According to Glivenko's theorem, $F_n(x)$ converges uniformly with probability 1 to $F(x)$ on the whole real axis.

Comparing this property with the parametric case, we wish to remark that the estimation of the distribution function given there does not lead to an unbiased estimator for all values of x. If, however, rather general conditions are satisfied, then this method is consistent. Estimating the unknown parameters from a large sample, we prefer the parametric estimation, because of the simplicity of the estimation procedure, the explicit functional form obtained, and the greater efficiency that can be achieved by it.

In order to estimate the density function in the nonparametric case, let us now consider the set of all distribution functions with bounded and uniformly continuous density function:

$$\left\{ F(x)\colon \frac{dF(x)}{dx} = f(x) \text{ uniformly continuous, bounded} \right\} . \tag{16}$$

Even for this set of distribution functions there does not exist an estimator for the density function that is unbiased for all x.

In 3.2.3.1 the density histogram belonging to a given subdivision $(a_0 < a_1 < \ldots < a_r)$ of an interval containing the sample elements was defined as

$$f_n(x) = \begin{cases} 0 & \text{if} \quad x < a_0 \\ \dfrac{g_i}{a_i - a_{i-1}}, & \text{if} \quad a_{i-1} < x < a_i, \quad i = 1, 2, \ldots r \\ 0 & \text{if} \quad a_r < x, \end{cases}$$

where g_i denotes the relative frequency of the interval (a_{i-1}, a_i) of the event $\{a_{i-1} < x < a_i\}$.

Here we can state the following: For the set of distribution functions given in (16), $f_n(x)$ tends stochastically to $f(x)$ as $n \to \infty$ and r tends to infinity with order of magnitude $\sqrt[3]{n}$ and $\max\limits_{(i)} (a_i - a_{i-1}) \to 0$, as $n \to \infty$.

(c) *Estimation of the expected value and the variance.* We summarize and extend the statements given above:

If the random variable X has a distribution with first and second moments μ_1 and μ_2 respectively, then the sample mean is an unbiased, strongly consistent estimator for the expected value $\mu_1 = E(x)$, and has minimum variance among all estimators which are linear functions of the sample elements:

$$\mu_1 \sim \overline{X}, \quad E(\overline{X}) = \mu_1, \quad D^2(\overline{X}) = \frac{D^2(X)}{n} = \frac{\mu_2 - \mu_1^2}{n}.$$

If, moreover, the fourth moment

$$\mu_4 = \int_{-\infty}^{\infty} x^4 f(x)\, dx$$

exists, then the empirical variance

$$s_n^2 = \frac{1}{n-1} \sum_{i-1}^{n} (X_i - \overline{X})^2$$

is an unbiased, strongly consistent estimator for $\sigma^2 = D^2(X)$:

$$\sigma^2 \sim s_n^2, \quad E(s_n^2) = \sigma^2, \quad D^2(s_n^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right).$$

4.2.3 Interval Estimation

4.2.3.1 The Notion of Confidence Interval. The Case of the Normal Distribution. Let X be distributed according to $N(\mu, \sigma_0)$ with known σ_0. As we know already, the best unbiased estimator for μ is the sample mean $\overline{X}$. We obtain an "interval estimate" for μ in the following way. The distribution of $\overline{X}$ is $N\left(\mu, \frac{\sigma_0}{\sqrt{n}}\right)$. Given the probability $1 - \varepsilon$ (with small ε) we can construct an interval $(-u_\varepsilon, u_\varepsilon)$ which contains the standardized variable $\frac{\overline{X} - \mu}{\sigma_0}\sqrt{n}$ with this probability:

$$P\left(-u_\varepsilon < \frac{\overline{X} - \mu}{\sigma_0}\sqrt{n} < u_\varepsilon\right) = 1 - \varepsilon. \tag{17}$$

We obtain u_ε by the relation

$$\Phi(u_\varepsilon) - \Phi(-u_\varepsilon) = 2\,\Phi(u_\varepsilon) - 1 = 1 - \varepsilon,$$

hence

$$\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}$$

E.g. if $\varepsilon = 0.045$, then $u_\varepsilon = 2$.

From (17) it follows that the following event has probability $1 - \varepsilon$:

$$\mu - u_\varepsilon \frac{\sigma_0}{\sqrt{n}} < \overline{X} < \mu + u_\varepsilon \frac{\sigma_0}{\sqrt{n}},$$

but from this we find that with the same probability the following relation holds:

$$\overline{X} - u_\varepsilon \frac{\sigma_0}{\sqrt{n}} < \mu < \overline{X} + u_\varepsilon \frac{\sigma_0}{\sqrt{n}}. \tag{18}$$

So we have obtained a random interval (α_1, α_2) with $\alpha_1 = \overline{X} - u_\varepsilon \frac{\sigma_0}{\sqrt{n}}$ and $\alpha_2 = \overline{X} + u_\varepsilon \frac{\sigma_0}{\sqrt{n}}$, which contains the unknown expected value with probability $1 - \varepsilon$. Applying this method many times with the same probability, $1 - \varepsilon$, the observed interval will contain, on the average, the true value of the parameter in $100\,(1 - \varepsilon)\,\%$ of the cases and will not contain it in $100\,\varepsilon\,\%$ of the cases.

The above method of interval estimation can be formulated in the following way:

Let X be a random variable, the distribution function of which depends on the parameter a. Given $1 - \varepsilon$ (ε small), if we succeed in constructing two statistics $\alpha_1 = \alpha_1(x_1, x_2, \ldots x_n;\ \varepsilon)$ and $\alpha_2 = \alpha_2(x_1, x_2, \ldots x_n;\ \varepsilon)$ such that

$$P(\alpha_1 < a < \alpha_2) = 1 - \varepsilon,$$

then we say that (α_1, α_2) is a *confidence interval* for the parameter a with *confidence level* $1 - \varepsilon$.

For the expected value of the Gaussian distribution the relation (18) gives a confidence interval with confidence level $1 - \varepsilon$.

If the distribution of X is different from the normal distribution but satisfies certain conditions, e.g., it has variance $D^2(X) = \sigma_0^2$, then the arithmetic mean $\bar{X}$ of a sample of size n taken on X is approximately distributed according to $N\left(E(X), \frac{\sigma_0}{\sqrt{n}}\right)$ for large n, and therefore we can write, to a good approximation

$$P\left(\bar{X} - u_\varepsilon \frac{\sigma_0}{\sqrt{n}} < E(X) < \bar{X} + u_\varepsilon \frac{\sigma_0}{\sqrt{n}}\right) \approx 1 - \varepsilon = 2\Phi(u_\varepsilon) - 1 .$$

For the theoretical variance $D^2(X) = \sigma^2$ of the normally distributed variable the construction of a confidence interval is as follows: the statistic $(n - 1)\frac{s_n^2}{\sigma^2}$ has a χ^2 distribution with parameter $n - 1$ (see 2.9.6); let us denote by $H_n(X)$ its distribution function and determine the value h_1 and h_2 with

$$P\left((n-1)\frac{s_n^2}{\sigma^2} < h_1\right) = H_n(h_1) = \frac{\varepsilon}{2},$$

$$P\left((n-1)\frac{s_n^2}{\sigma^2} < h_2\right) = H_n(h_2) = 1 - \frac{\varepsilon}{2},$$

i.e., the $\frac{\varepsilon}{2}$ and $1 - \frac{\varepsilon}{2}$ quantiles.

From this we obtain the relation

$$P\left(h_1 < (n-1)\ \frac{s_n^2}{\sigma^2} < h_2\right) = 1 - \varepsilon;$$

hence

$$\sqrt{n-1}\,\frac{s_n}{\sqrt{h_2}} < \sigma < \sqrt{n-1}\,\frac{s_n}{\sqrt{h_1}}$$

will be the confidence interval for the standard deviation with confidence level $1 - \varepsilon$.

4.2.3.2 Confidence Interval for the Probability. Let A be an event with probability $p = P(A)$; let us denote by $\varkappa_n$ the random variable which is the frequency of A in n observations; further let $\gamma_n = \frac{\varkappa_n}{n}$ be the relative frequency. We now make a distinction between the cases when n is large and when it is not.

If n is large then the distribution of $\frac{\gamma_n - p}{\sqrt{np(1-p)}}$ is close to the standard normal distribution. Thus the approximate relation

$$P\left(-u_\varepsilon < \frac{\gamma_n - p}{\sqrt{np(1-p}} < u_\varepsilon\right) \approx 1 - \varepsilon = 2\,\Phi(\varepsilon) - 1$$

holds, from which we obtain after some calculations, for the confidence interval of p with a level close to $1 - \varepsilon$:

$$\frac{\gamma_n + \frac{u_\varepsilon^2}{2n} - \frac{u_\varepsilon}{\sqrt{n}}\sqrt{\gamma_n(1-\gamma_n) + \frac{u_\varepsilon^2}{4n}}}{1 + \frac{u_\varepsilon^2}{n}} < p < \frac{\gamma_n + \frac{u_\varepsilon^2}{2n} + \frac{u_\varepsilon}{\sqrt{n}}\sqrt{\gamma_n(1-\gamma_n) + \frac{u_\varepsilon^2}{4n}}}{1 + \frac{u_\varepsilon^2}{n}}\,.$$

If n is large enough so that $\frac{1}{n}$ is negligible compared to $\frac{1}{\sqrt{n}}$, then we can simplify our expression:

$$\gamma_n - \frac{u_\varepsilon}{\sqrt{n}}\sqrt{\gamma_n(1-\gamma_n)} < p < \gamma_n + \frac{u_\varepsilon}{\sqrt{n}}\sqrt{\gamma_n(1-\gamma_n)}\,.$$

If n is a small integer, then for the construction of a confidence interval we have to take into account the exact probability distribution of $\varkappa_n$, i.e., the binomial distribution. Omitting details, we give the method of obtaining the confidence interval: For every $p (0 < p < 1)$ we determine two constants k_1 and k_2 which satisfy the relations

$$\sum_{i=0}^{k_1} \binom{n}{i} p^i (1-p)^{n-i} < \frac{\varepsilon}{2}\,, \quad \text{but} \quad \sum_{i=0}^{k_1+1} \binom{n}{i} p^i (1-p)^{n-i} \geqslant \frac{\varepsilon}{2}\,,$$

$$\sum_{i=k_2}^{n} \binom{n}{i} p^i (1-p)^{n-i} < \frac{\varepsilon}{2}, \quad \text{but} \quad \sum_{i=k_2-1}^{n} \binom{n}{i} p^i (1-p)^{n-1} \geqslant \frac{\varepsilon}{2}$$

respectively. We establish a coordinate system (p, k) where we mark both values k_1 and k_2 corresponding to each value of p; in this way we obtain two step functions (see Fig. 22). Carrying out observations and getting the result $\varkappa_n = k$, we draw a straight line of height k parallel to the p axis. This line intersects the mentioned step functions at points belonging to values p_1 and p_2 $(p_1 < p_2)$; these values p_1 and p_2 give the confidence interval for p with the level $1 - \varepsilon$: $p_1 < p < p_2$.

A similar method leads to the confidence interval for the parameter of the Poisson distribution.

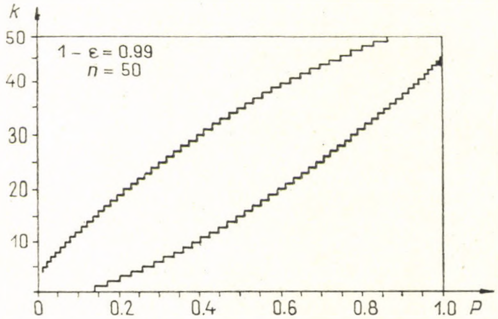


FIG. 22. Construction of a confidence interval for the parameter p of the binomial distribution at confidence level 0.99

4.2.3.3 Confidence Interval for the Parameter of the Exponential Distribution. Let X be distributed according to the exponential law:

$$P(X \leqslant x) = 1 - \varepsilon^{-\lambda x}, \quad 0 < \lambda < \infty, \quad x \geqslant 0.$$

As mentioned before (see 2.9.5), the sum $n\overline{X} = X_1 + X_2 + \ldots + X_n$ follows a gamma distribution of order n; denoting this distribution function by $G_n(x; \lambda)$ we may write

$$P(n\overline{X} \leqslant x) = G_n(x; \lambda) = \int_0^x \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}\, dt, \; x \geqslant 0.$$

As for the exponential variable

$$E(X) = \frac{1}{\lambda} \quad \text{and} \quad E(\lambda X) = 1;$$

the random variable $n\bar{X}\lambda$ has the distribution function $G_n(x; 1)$. This is fully determined and has expected value n and standard deviation $\sqrt{n}$. Given $1 - \varepsilon$, we determine the values $h_n\left(\frac{\varepsilon}{2}\right)$ and $h_n\left(1 - \frac{\varepsilon}{2}\right)$ in such a way that

$$G_n\left(h_n\left(\frac{\varepsilon}{2}\right); 1\right) = \frac{\varepsilon}{2} \text{ and } G_n\left(h_n\left(1 - \frac{\varepsilon}{2}\right); 1\right) = 1 - \frac{\varepsilon}{2}.$$

Then with probability $1 - \varepsilon$ the relation

$$h_n\left(\frac{\varepsilon}{2}\right) \leqslant n\bar{X}\lambda \leqslant h_n\left(1 - \frac{\varepsilon}{2}\right)$$

holds, from which we obtain for λ the confidence interval with level $1 - \varepsilon$:

$$\frac{1}{n\bar{X}} h_n\left(\frac{\varepsilon}{2}\right) \leqslant \lambda \leqslant \frac{1}{n\bar{X}} h_n\left(1 - \frac{\varepsilon}{2}\right).$$

Since $2n\lambda\bar{X}$ is distributed according to the χ^2 distribution with parameter $2n$ (Section 2.9.6), the constants $2h_n\left(\frac{\varepsilon}{2}\right)$ and $2h_n\left(1 - \frac{\varepsilon}{2}\right)$ can be determined from the χ^2 table [C4, C5, C9].

4.2.3.4 Some Important Remarks Concerning Confidence Intervals. Considering the confidence interval for the expected value of a normal variable, the relation

$$\bar{X} - u_\varepsilon \frac{\sigma_0}{\sqrt{n}} < \mu < \bar{X} + u_\varepsilon \frac{\sigma_0}{\sqrt{n}} \tag{19}$$

holds with probability $1 - \varepsilon$. This is true in the sense given above, the confidence interval in (19) being a *random interval.*

Let us assume now that we have carried out the sampling and have obtained for $\bar{X}$ the observed value $\bar{x}$. In this case we cannot say that the interval $\left(\bar{x} - u_\varepsilon \frac{\sigma_0}{\sqrt{n}}, \bar{x} + u_\varepsilon \frac{\sigma_0}{\sqrt{n}}\right)$ contains the unknown parameter μ with probability $1 - \varepsilon$: we have now no random interval, but a numerically given interval which either contains μ or does not.

As a second remark we consider the length of the interval. In the Gaussian case the confidence interval for μ has the length $2u_\varepsilon \frac{\sigma_0}{\sqrt{n}}$. E.g., if $1 - \varepsilon = 0.9973$, $u_\varepsilon = 3$ and we have $6 \frac{\sigma_0}{\sqrt{n}}$. If this is not small enough for practical purposes we can go in two different directions. We can diminish the value of the confidence level $1 - \varepsilon$; in this case the average number of cases when the interval does not contain the true value of the parameter will increase. If, e.g., $1 - \varepsilon = 0.9453$ then $u_\varepsilon = 2$ and the length of the interval will be $4 \frac{\sigma_0}{\sqrt{n}}$ which is 2/3 of the earlier one; at the same time the average number of wrong decisions increases from 0.27% to 5.47% of the cases.

A second possibility is to increase the sample size; the length of the interval in the treated example is inversely proportional to the number of observations.

For a given error e, or relative error e_0, the relations

$$2\, u_\varepsilon \frac{\sigma_0}{\sqrt{n}} < e, \qquad \frac{2\, u_\varepsilon}{\bar{x}} \frac{\sigma_0}{\sqrt{n}} < e_0 .$$

may be used to determine an appropriate choice of the confidence level and sample size.

Our final remark concerns the location of the confidence interval. In our examples two different methods were used. If the distribution of the statistic under consideration was symmetric about the parameter in question, symmetric intervals were chosen. In such cases, this leads to the shortest intervals in most practical situations. If the distribution is nonsymmetric, the determination of a shortest interval is not at all simple in general. In our examples we have chosen the $\frac{\varepsilon}{2}$ and $1 - \frac{\varepsilon}{2}$ quantiles of the distribution of the statistic.

In certain cases, in spite of the asymmetry of the distribution, symmetric intervals are used. This is not always suitable, as in the case of a positive variable where the lower limit may happen to turn out negative.

For large sample sizes, in many cases, the distribution of the statistic is close to the normal distribution in the sense that the standardized variable $\frac{\alpha_n - E(\alpha_n)}{D(\alpha_n)}$ tends in distribution to the Gaussian law. Therefore the normal distribution can be applied in such cases for large values of n.

4.2.4 The Method of Bayes

4.2.4.1 Introduction. One of the earliest results in probability theory, due to T. H. Bayes (✝ 1761), was published in his posthumus article "An essay towards solving a problem in the doctrine of chances". His theorem, which we have treated in Sections 1.9 and 2.11, enables us to determine, under certain conditions, the distribution function of a random variable by means of earlier experiences; that is why this theorem is an important tool in mathematical statistics. We will turn now to solving the problem raised in Sections 1.9 and 2.11.

We turn to a slightly modified form of Experiment 4'. From time to time the factory delivers lots of products, consisting of (say) 10,000 items each, to the motor works where the parts will be mounted. Let us suppose that in the course of this routine the defectives are discovered and isolated; in this way the fraction defective of each lot can subsequently be stated. Let us denote by π the fraction defective in a lot — a random variable; contrary to Experiment 4' in Section 1.9 we now consider this quantity as a continuous random variable instead of a discrete one. We can determine its approximate distribution function from data collected in a long run of deliveries. Let

$$P(\pi \leqslant x) = W(x).$$

If we have to make some statement about the fraction defective in a lot just obtained, we can only say that it must be with large probability (say $1 - \varepsilon$) between some two values p_1 and p_2 ($p_1 < p_2$) for which

$$P(p_1 \leqslant \pi \leqslant p_2) = 1 - \varepsilon$$

where we can choose p_1 and p_2 so that the following relations hold:

$$W(p_1) = \frac{\varepsilon}{2}, \qquad W(p_2) = 1 - \frac{\varepsilon}{2}.$$

Let us suppose, on the other hand, that we have taken from the present lot a random sample of size n and have found k defective pieces in it. If the random variable $\varkappa_n$ denotes the number of defective items in a sample of size n then we may ask: what is the distribution function of π under the condition that the event $\varkappa_n = k$ occurred:

$$P(\pi \leqslant x \mid \varkappa_n = k) = W(x|k).$$

If we succeed in making use of both the previous information concerning these lots in general, and the present information concerning our lot in question, we should acquire better knowledge of the quality of the actual lot; e.g., a shorter interval than (p_1, p_2) for the same level $1 - \varepsilon$, etc.

The problem of determining the posterior distribution $P(\pi \leqslant x \mid k_n = k)$ slightly differs from the cases treated in Sections 1.9 and 2.11

because here one of the variables is continuous while the other is discrete. This change, however, means no real difficulty: the version of Bayes' theorem for this case can be deduced in the same way (see Section 4.2.4.4)

In the next Section we will discuss the problem in the case when we do not have complete information about $W(x)$. Finally, we will consider the mentioned example and an additional one.

4.2.4.2 The Method. The theorem applies if we know the prior distribution of the random parameter. In many situations, however, this is unknown, or we have only a partial knowledge of it, either from theoretical considerations or from statistical data. Bayes himself considered the question only for the finite, discrete case and suggested that the prior distribution be assumed to be uniform: i.e., if ξ is a random variable with possible values $1, 2, \ldots, m$, then the prior distribution is

$$P(\xi = i) = \frac{1}{m}, \quad i = 1, 2, \ldots, m.$$

This assumption, called the Bayesian method, gave rise to a long and heated debate among mathematicians over whether it could be regarded as justified. After many years, the following conclusions were drawn: Having no knowledge of the prior distribution, we may start with any hypothetical distribution function (if it seems to be in accordance with the situation and if it is not too laborious) and determine the posterior distribution. This result has to be compared with subsequent observations, and, if the agreement is good, the correctness of the hypotheses will be accepted. Otherwise the hypotheses will have to be changed.

There was, however, a further argument in favor of the uniform distribution: if the observed event, i.e. the condition, is based on n independent observations and n is sufficiently large, then under certain conditions the posterior distribution depends only to a small extent on the prior distribution. This is the case in the two examples to follow. The proof is based on a simple lemma of Laplace which is frequently used in statistical theory and practice. We are first going to consider this lemma.

4.2.4.3 The Lemma of Laplace. Let $h(x)$ be continuous in the interval (a, b) $(-\infty \leqslant a < b \leqslant \infty)$ with the following properties:

$h(c) = 1$ for a value c with $a < c < b$,

$h(x) < 1$ for $x \neq c$,

$h'(x)$ and $h''(x)$ are continuous in $(c - \delta,\ c + \delta)$ for some $\delta > 0$,

$$h'(c) = 0,\quad h''(c) = -d,\quad d > 0.$$

Further let $g(x)$ be continuous in (a, b) and

$$\int_a^b h(t)\, g(t)\, dt < \infty$$

Then, the following relations hold:

Lemma of Laplace.

With $x > 0$, $y > 0$

$$\lim_{n\to\infty} \sqrt{\frac{nd}{2\pi}} \int_{c-\frac{x}{\sqrt{nd}}}^{c+\frac{y}{\sqrt{nd}}} h^n(t)\, g(t)\, dt = g(c)\left[\Phi(y) - \Phi(-x)\right]$$

and

$$\lim_{n\to\infty} \sqrt{\frac{nd}{2\pi}} \int_a^b h^n(t)\, g(t)\, dt = g(c)$$

where $\Phi(x)$ denotes the distribution function of the standard normal law.

This means, e.g., that for sufficiently large n the following asymptotic relations hold:

$$\int_{c-\frac{x}{\sqrt{nd}}}^{c+\frac{y}{\sqrt{nd}}} h^n(t)\, g(t)\, dt \approx \sqrt{\frac{2\pi}{nd}}\, g(c)\left[\Phi(y) - \Phi(-x)\right],$$

$$\int_a^b h^n(t)\, g(t)\, dt \approx g(c) \sqrt{\frac{2\pi}{nd}},$$

i.e., as $n \to \infty$, the left-hand terms tend to zero in order of magnitude $\frac{1}{\sqrt{n}}$.

4.2.4.4 Two Examples of the Application of the Bayesian Method. First example: We turn now to the problem mentioned in the introduction, using the notations given there. Let the prior density function of the fraction defective π be the continuous function $w(x)$. Under the condition that the sample of size n contained $\varkappa_n = k$ defective items, the posterior density of π is, according to the Bayes' theorem, the following:

$$w(x \mid k) = \frac{P(\varkappa_n = k \mid \pi = x)\, w(x)}{\int\limits_0^1 P(\varkappa_n = k \mid \pi = t)\, w(t)\, dt}$$

which is formally the same as both forms of the theorem given in Sections 1.9 and 2.11. Here $P(\varkappa_n = k \mid \pi = x)$ denotes the probability that a sample of size n contains k defective items if the fraction defective is x. Under this condition, however, $\varkappa_n$ is distributed according to the binomial law, i.e.,

$$P(\varkappa_n = k \mid \pi = x) = \binom{n}{k} x^k (1-x)^{n-k}; \qquad k = 0, 1, 2, \ldots, n.$$

Finally

$$w(x \mid k) = \frac{x^k (1-x)^{n-k}\, w(x)}{\int\limits_0^1 t^k (1-t)^{n-k}\, w(t)\, dt}; \quad 0 \leqslant x \leqslant 1. \tag{20}$$

Let us now consider two cases:

(a) Let $w(x)$ be unknown and n large, with $n \to \infty$ such that the ratio of defective items in the sample $\frac{k}{n} \to q$ $(0 < q < 1)$, or, for the sake of simplicity, let $k = [qn]$.

In this case it is easy to show that the functions

$$h(x) = \frac{x^q (1-x)^{1-q}}{q^q (1-q)^{1-q}} \text{ and } g(x) = w(x) = W'(x)$$

satisfy the conditions of the lemma of Laplace with $a = 0$, $b = 1$. By using the above notation, from elementary calculations we obtain

$$c = q, \quad d = \frac{1}{q(1-q)}$$

$$P\left(x_1 < \pi < x_2 \,\middle|\, \frac{\varkappa_n}{n} = q\right) = W\left(x_2 \,\middle|\, \frac{\varkappa_n}{n} = q\right) - W\left(x_1 \,\middle|\, \frac{\varkappa_n}{n} = q\right)$$

$$= \frac{\int\limits_{x_1}^{x_2} h^n(t)\, g(t)\, dt}{\int\limits_0^1 h^n(t)\, g(t)\, dt}, \quad 0 \leqslant x \leqslant 1.$$

For the denominator the lemma gives directly

$$\int_0^1 h^n(t)\, g(t)\, dt \approx g(c) \sqrt{\frac{2\pi}{nd}} = w(q) \sqrt{2\pi \frac{q(1-q)}{n}} .$$

In the numerator we place x_1 and x_2 near q in the following sense:

$$x_1 = q - y_1 \sqrt{\frac{q(1-q)}{n}}, \quad x_2 = q + y_2 \sqrt{\frac{q(1-q)}{n}}, \quad y_1 > 0,\ y_2 > 0 .$$

So we have

$$\int\limits_{q - \frac{y_1}{\sqrt{n \frac{1}{q(1-q)}}}}^{q + \frac{y_2}{\sqrt{n \frac{1}{q(1-q)}}}} h^n(t)\, q(t)\, dt \approx g(c) \sqrt{\frac{2\pi}{nd}} \left[\Phi(y_2) - \Phi(-y_1)\right] =$$

$$= \sqrt{\frac{2\pi q(1-q)}{n}} \left[\Phi\left(\frac{x_2 - q}{\sqrt{\frac{q(1-q)}{n}}} \right) - \Phi\left(\frac{x_1 - q}{\sqrt{\frac{q(1-q)}{n}}} \right) \right] w(q) \qquad (21)$$

and (21) result in

$$W\left(x \,\middle|\, \frac{\varkappa_n}{n} = q\right) \approx \Phi\left(\frac{x - q}{\sqrt{\frac{q(1-q)}{n}}} \right).$$

This means that if n is sufficiently large, then the posterior distribution of the fraction defective is asymptotically normal with mean q and variance $\sqrt{\frac{q(1-q)}{n}}$. For example, if $n = 200$, $k = 8$, then $q = 0.04$ and

$$P\left(q - 2\sqrt{\frac{q(1-q)}{n}} < \pi < q + 2\sqrt{\frac{q(1-q)}{n}} \,\middle|\, q\right) \approx 0\,95$$

i.e., $P(0.0123 < \pi < 0.0677 \mid q = 0.04) = 0.95 = 95\%$.

(b) Under certain conditions the fluctuation of the fraction defective can be described by means of a beta distribution (see Section 2.9.9) with two parameters (let us denote them by r and s), with a density function as follows:

$$w(x) = B(r, s) x^{r-1} (1 - x)^{s-1}, \quad 0 \leqq x \leqq 1,$$

where

$$B(r,s) = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}\,.$$

Let us suppose that the prior distribution function of π is $W(x) = \int\limits_0^x w(t)dt$, $0 \leqslant x \leqslant 1$. The mean and variance of π are the following:

$$E(\pi) = \frac{r}{r+s}\,, \quad D^2(\pi) = \frac{rs}{(r+s)^2\,(r+s+1)}\,.$$

The posterior density of π, given $\varkappa_n = k$, can be obtained by substituting $w(x)$ into (20):

$$w(x\,|\,k) = \frac{x^{k+r-1}\,(1-x)^{n-k+s-1}}{\int\limits_0^1 t^{k+r-1}\,(1-t)^{n-k+s-1}\,dt} =$$
$$= B(k+r, n-k+s)\,x^{k+r-1}\,(1-x)^{n-k+s-1}, \quad 0 \leqslant x \leqslant 1.$$

(The validity of the last formula can be seen from Section 2.9.9.) This means that the posterior distribution of π, given $\varkappa_n = k$, is a beta distribution as well, but with parameters $r+k$ and $s+n-k$, resp. The mean value and variance are in this case

$$E(\pi\,|\,\varkappa_n = k) = \frac{k+r}{n+r+s}\,,$$

$$D^2(\pi\,|\,\varkappa_n = k) = \frac{(r+k)\,(n-k+s)}{(n+r+s)^2\,(x+r+s+1)}\,.$$

E.g. if $r = 3$, $s = 47$, $n = 200$, $k = 8$

$E(\pi) = 0.06 = 6\ \%$, $D(\pi) = 0.0333$

then $E(\pi|k) = 0.044 = 4.4\%$, $D(\pi|k) = 0.013$.

Second example: The quality of tomato purée is usually characterized by its so-called Howard count, which can be obtained in the following way: A sample is diluted according to standard prescriptions and from the solution some drops are examined under the microscope in a Burker chamber. The number of cells (out of one hundred) containing mould figures is called the Howard count of the sample. Its distribution for a given lot is close to the normal law. Using a day's production as a lot (or the production of a constant or variable time interval, depending on the raw material etc.), the expected value of the Howard count is a random variable, fluctuating from lot to lot.

Let us denote the Howard count determined from a random sample by η and let $X = E(\eta)$ be its expectation within the lot. X itself is a random variable with (prior) distribution

$$P(X \leqslant x) = F(x) = \int_{-\infty}^{x} f(t)dt, \quad -\infty < x < \infty.$$

Suppose now that in n independent samples taken from the lot, the Howard count η has taken on the values $\eta_1 = y_1, \eta_2 = y_2, \ldots \eta_n = y_n$, with arithmetic mean $\bar{\eta} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$. What is the posterior distribution of X given $\bar{\eta} = \bar{y}$? For the density function we obtained, according to Section 2.11, the following expression:

$$f(x \mid \bar{y}) = \frac{f(\bar{y} \mid x)\, f(x)}{\int_{-\infty}^{\infty} f(\bar{y} \mid t)\, f(t)\, dt}, \quad -\infty < x < \infty.$$

$f(\bar{y} \mid x)$ is the density function of the sample mean under the condition $X = x$; as mentioned the corresponding conditional distribution is $N\left(x, \frac{\sigma^2}{n}\right)$, i. e.

$$f(\bar{y} \mid x) = \varphi\left(\bar{y}; x, \frac{\sigma^2}{n}\right) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} e^{\frac{n(\bar{y}-x)^2}{2\sigma^2}}, \quad -\infty < \bar{y} < \infty.$$

Finally, we have

$$f(x \mid \bar{y}) = \frac{e^{-\frac{n(\bar{y}-x)^2}{2\sigma^2}} f(x)}{\int_{-\infty}^{\infty} e^{-\frac{n(\bar{y}-t)^2}{2\sigma^2}} f(t)\, dt}.$$

Using the lemma of Laplace with

$$h(x) = e^{-\frac{(\bar{y}-x)^2}{2\sigma^2}}, \quad g(x) = f(x),$$

$$a = -\infty, \quad b = +\infty, \quad c = \bar{y}, \quad d = \frac{1}{\sigma^2},$$

we obtain for large n the (asymptotic) posterior distribution

$$F(x \mid \bar{y}) \equiv \Phi\left(x; \bar{y}, \frac{\sigma^2}{n}\right) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{n(\bar{y}-t)^2}{2\sigma^2}} dt, \quad -\infty < x < \infty$$

i.e., if $\overline{\eta} = \overline{y}$, then the expected value of the Howard count is distributed according to the normal law, with mean $\overline{y}$ and variance $\frac{\sigma^2}{n}$ ·

If $f(t) = \varphi(t;\ \mu,\ \sigma)$, i.e., if the prior density is Gaussian, the Bayesian method can be easily applied and evaluated.

4.3 TESTING STATISTICAL HYPOTHESES

4.3.1 The Statistical Test Procedure

4.3.1.1 Statistical Hypotheses. We begin by presenting some examples. Similar problems are often met in the practice of quality control and in industrial experimentation. The method for answering these questions, the so-called *statistical test procedure,* is one of the main tools in statistical quality control. We turn now to our examples:

(a) Lots of cogwheels are delivered steadily by a factory, each lot containing, eg., $N = 5{,}000$ pieces. We are able to ascertain if a piece is good or defective. Let us assume that a lot is considered as a good one if its fraction defective does not exceed the value $p = 0.02 = 2\%$; otherwise it is a bad lot and is to be rejected. How can we decide to accept or reject on the basis of a sample?

(b) A machine produces cylinders. From earlier experience (from a large sample) we know that the length X of a single piece, as a random variable, was distributed according to $N(\mu_0 = 23\text{ mm},\ \sigma_0 = 0.15\text{ mm})$. Our present aim is to examine whether the fluctuation is still centered at the expected value $\mu_0 = 23$ mm or not. We assume that a change in the standard deviation occurs very rarely, and that this usually takes place together with a change in the expected value. To answer the question we take a relatively small sample from the production of the last hour.

(c) Cylinders are produced on two machines M_1 and M_2 of the same type. The products may be treated as a homogeneous lot if the distributions of the length, diameter and breaking strength are the same in the case of both machines. How can we convince ourselves of the agreement of the distributions by means of samples taken from the machines?

In the above problems one or more random variables occur and questions are raised concerning their distribution functions or certain parameters of the distributions. The problem can be formulated in each case in the following way: A statement or assumption is given concerning the distribution and a decision is to be made as to whether this statement should be accepted or rejected. The decision has to be based on samples.

As the answer is influenced by chance, our decision may happen to be wrong. Let us formulate the problems more accurately:

(a) Having a lot containing N pieces we take a random sample of size n from it. Denoting by $\varkappa_n$ the number of defective items in the sample, $\varkappa_n$ is distributed according to the binomial law with unknown parameter

$$E\left(\frac{\varkappa_n}{n}\right) = p, \qquad 0 \leqslant p \leqslant 1.$$

We have now to decide whether the statement

$$H_0\colon\ p \leqslant p_0 = 0.02$$

can be accepted or the opposite statement

$$H_1\colon\ p > p_0 = 0.02$$

has to be accepted. The decision will be based on the observed value k of the random variable $\varkappa_n$.

We have denoted here by H_0 the statement that the lot is a good one: H_0 is called the *null hypothesis*. The statement H_1 is called the *alternative hypothesis*. The test will result either in the acceptance of the null hypothesis H_0 or in the rejection of it (acceptance of the alternative hypothesis H_1). In statistical work usually the standard, or the favorable case, is taken as null situation hypothesis. In our example the assumption that the lot is a good one would be taken as H_0.

We turn now to the formulation of the other two problems using the same terminology:

(b) According to our assumptions the length X of the cylinder is distributed according to the normal law with given standard deviation:

$$P(X \leqslant x) = \Phi(x;\ \mu,\ \sigma_0), \qquad \sigma_0 \text{ known.}$$

This statement is part of both hypotheses.

The standard situation will be taken as null hypothesis:

$$H_0\colon\ \mu = 23.00 \text{ mm}, \qquad \sigma_0 = 0.15 \text{ mm},$$

while for the alternative hypothesis we have:

$$H_1\colon\ \mu \neq \mu_0 = 23.00 \text{ mm}, \qquad \sigma_0 = 0.15 \text{ mm}$$

(c) Of the three mentioned quantities let us consider the breaking strength. We denote it by $X^{(1)}$ if produced by the machine M_1 and by $X^{(2)}$ if produced by the machine M_2. Each of them has a certain distribution:

$$P(X^{(1)} \leqslant x) = F_1(x), \quad P(X^{(2)} \leqslant x) = F_2(x).$$

In the favorable case, these two distributions agree; accordingly, the null hypothesis is chosen as:

$$H_0\colon\ F_1(x) \equiv F_2(x),$$

while the alternative hypothesis will be

$$H_1\colon\ F_1(x) \not\equiv F_2(x).$$

Let us draw our attention to the fact that in our examples two different kinds of statistical hypotheses occur: the distribution function of the random variable is either completely determined by the hypothesis, or the hypothesis contains a set of distribution functions. The first case occurs in (b) where under H_0 the length of the cylinder has the distribution function $\Phi(x;\ \mu_0 = 23,\ \sigma_0 = 0.15)$. In this case we speak of *simple hypothesis*. In all other cases we have *composite hypotheses;* they contain a set of distributions.

This set is, e.g., in H_0 of (a) all binomial distributions with parameter $p\colon\ 0 \leqslant p \leqslant p_0 = 0.02$; in H_1 of (b) all Gaussian distribution functions with standard deviation $\sigma_0 = 0.15$ and mean value different from $\mu_0 = 23$:

$$\{\Phi(x;\ \mu,\ \sigma_0)\colon\ \mu \neq \mu_0 = 23,\quad \sigma_0 = 0.15\}.$$

The situation is a little different in H_0 of (c). This seems to be at first sight a simple hypothesis, but what really happens is the following. We do not know in advance (as in many cases) the type of the distribution of the random variable; in other cases these types may change from one production process or raw material to the other. The latter may happen for the breaking strength, where the only thing we know is that the distribution function is a continuous one. Now we can characterise the null hypothesis as the set of all pairs of continuous distribution functions, the two elements of which agree:

$$H_0\colon\ \{(F_1(x);\ F_2(x))\colon\ F_1(x) \equiv F_2(x);\ F_1(x),\ F_2(x)\ \text{continuous}\},$$

while the alternative is

$$H_1\colon\ \{(F_1(x),\ F_2(x))\colon\ F_1(x) \not\equiv F_2(x);\ F_1(x),\ F_2(x)\ \text{continuous}\}.$$

In the cases (a) and (b) both H_0 and H_1 contain parametric sets of distribution functions (binomial, normal) but we cannot say that in case (c). Hence, the latter is a nonparametric problem.

4.3.1.2 The Test Procedure. Our aim is to make a decision between two hypotheses; our decision must be based on observations concerning the random variable or variables concerned. In order to sketch the decision procedure, we turn to problem (b). In this example, concerning the distribution of the length X, the null hypothesis

$$H_0\colon P(X \leqslant x) = \Phi(x;\ \mu_0,\ \sigma_0),\ \mu_0 = 23;\ \sigma_0 = 0.15,$$

is to be tested against the alternative

$$H_1\colon P(X \leqslant x) = \Phi(x;\ \mu,\ \sigma_0),\ \mu \neq \mu_0 = 23,\ \sigma_0 = 0.15.$$

Let us take n independent observations on X: $(X_1, X_2, \ldots, X_n)$. The result is a point of the n dimensional Euclidean space, i.e., of the sample space. The test procedure or the *statistical test* means essentially a division of the sample space into two parts: in one of them the hypothesis H_0 will be rejected, while in the other it will be accepted. Accordingly, we have in the sample space a *rejection region* or *critical region*, denoted by $\mathfrak{X}_r$, and an acceptance region, denoted by $\mathfrak{X}_a$. This division is performed by means of a suitably chosen statistic in the following way: We know that in the case of a Gaussian variable a good estimator for the expected value is given by the sample mean

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

In the case of the validity of the null hypothesis the statistic $\overline{X}$ has distribution $N\left(\mu_0, \frac{\sigma_0}{\sqrt{n}}\right)$. If we now prescribe the probability $1 - \alpha$ near to 1 (i.e., α small) then we can determine the value u_a in such a way that for the Gaussian variable $\overline{X}$ the following relation holds:

$$P\left(\mu_0 - u_\alpha \frac{\sigma_c}{\sqrt{n}} < \overline{X} < \mu_0 + u_\alpha \frac{\sigma_0}{\sqrt{n}} \,\middle|\, H_0\right) = 1 - \alpha,$$

where u_a is determined by the following relation:

$$\Phi(u_\alpha) - \Phi(-u_\alpha) = 1 - \alpha, \text{ or } 2\Phi(u_\alpha) - 1 = 1 - \alpha.$$

It is convenient to introduce the standardized variable

$$u = \frac{\overline{X} - \mu_0}{\sigma_0/\sqrt{n}},$$

which in the case of the validity of H_0 is distributed according to $N(0.1)$. If we now require probability $1 - \alpha$ to be near to 1, e.g., $1 - \alpha = 0.99$

($\alpha = 0.01$) then we can determine the value $u_\alpha = 2.58$. This means that the sample mean $\overline{X}$ lies in the neighborhood

$$\left(\mu_0 - 2.58 \frac{\sigma_0}{\sqrt{n}}, \quad \mu_0 + 2.58 \frac{\sigma_0}{\sqrt{n}}\right)$$

of the expected value μ_0 with probability $1 - \alpha = 0.99$; or in other words the statistic u lies in the interval $(-2.58;\ 2.58)$ with probability 0.99:

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-2.58}^{2.58} e^{-\frac{t^2}{2}} dt = \Phi(2.58) - \Phi(-2.58) = 0.99 .$$

Hence, if the hypothesis H_0 is valid, then X will lie near the expected value with a probability close to 1, and there remains only a small probability ($\alpha = 1 - 0.99 = 0.01 = 1\%$) of its being out of the mentioned interval. This fact is quite natural and justifies the following test procedure:

We accept H_0 if a pre-assigned event with high probability occurs, i.e., if the sample $(X_1, X_2, \ldots, X_n)$ satisfies the relation:

$$\mu_0 - u_\alpha \frac{\sigma_0}{\sqrt{n}} < \overline{X} = \frac{X_1 + X_2 + \ldots X_n}{n} < \mu_0 + u_\alpha \frac{\sigma_0}{\sqrt{n}}$$

or

$$-u_\alpha = -\ 2.58 < u < u_\alpha = 2.58;$$

while we reject H_0 if the complementary event with small probability takes place, i.e., if one of the following two relations holds:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} < \mu_0 - u_\alpha \frac{\sigma_0}{\sqrt{n}} = c' .$$

or

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} > \mu_0 + u_\alpha \frac{\sigma_0}{\sqrt{n}} = c''$$

i.e., if either $u < -\ 2.58$, or $u > 2.58$.

The sample space is therefore divided into the critical region

$$\mathfrak{X}_c = \{(X_1, X_2, \ldots X_n)\colon X_1 + X_2 + \ldots + X_n < nc' \text{ or } \\ X_1 + X_2 + \ldots + X_n > nc''\}$$

and the acceptance region

$$\mathfrak{X}_a = \{(X_1, X_2, \ldots X_n)\colon nc' < X_1 + X_2 + \ldots + X_n < nc''\}.$$

Substituting the numerical values $\mu_0 = 23, \sigma_0 = 0.15$ and taking the sample size $n = 9$, we obtain for the constants

$$c' = \mu_0 - \mu_\alpha \frac{\sigma_0}{\sqrt{n}} = 22.871, \quad c'' = \mu_0 + u_\alpha \frac{\sigma_0}{\sqrt{n}} = 23.129.$$

Figure 23 illustrates the character of the critical region in the case of testing the sample mean for $n = 2$. For $n = 3$ or $n > 3$ two parallel planes and hyperplanes, respectively, will take over the role of the two boundary lines.

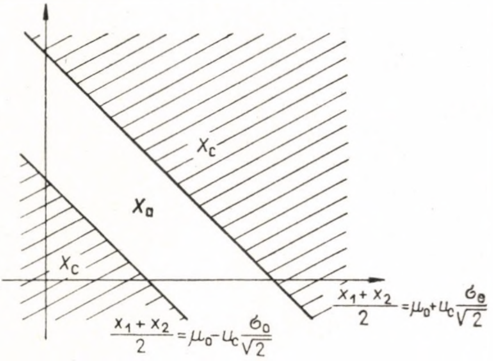

X_c
X_0
X_c
$\frac{x_1 + x_2}{2} = \mu_0 + u_c \frac{\sigma_0}{\sqrt{2}}$
$\frac{x_1 + x_2}{2} = \mu_0 - u_c \frac{\sigma_0}{\sqrt{2}}$

FIG. 23. Critical and acceptance regions in the sample space $R_2 = \{(X_1, X_2)\}$

As in the performance of the test procedure, the decision is based only on the function $\bar{X}$ of the sample values $(X_1, X_2, \ldots, X_n)$. The critical region and acceptance region will be used with respect to this statistic in the following sense: With the aid of the statistic $\bar{X}$, the n-dimensional sample space is mapped into the real line $(-\infty, \infty)$, where $\bar{X}$ takes on its possible values; this is the sample space belonging to the random variable $\bar{X}$. Here the critical region consists of the points of the interval $(-\infty, c')$ and (c'', ∞), while the acceptance region is the interval $(c'; c'')$. The situation is essentially the same in the case of other tests where the procedures are also based on some sample statistics.

Important questions soon to be treated are the following: how should we select an appropriate test statistic? what do we mean by saying that our choice is a most favorable one? etc. We mention here that if a sufficient statistic exists then it is preferred from the point of view of the information obtained from the sample concerning the true value of the parameter. But sometimes economic considerations may lead to the choice of a nonsufficient statistic; this happens, e.g., if the numerical determination of the latter is much simpler, affords less time and cost.

The probability $\Phi(2.58) - \Phi(-2.58) = 0.99$ in our example is called the *level of the test* or the *probability level*, while the complementary value $1 - \Phi(2.58) + \Phi(-2.58) = 2(1 - \Phi(2.58)) = 0.01$ is the *size of the test* or the significance level.

The structure of the test procedure is essentially the same in other cases as well: A statistic, i.e., a function $T = T(X_1, X_2, \dots X_n)$ of the sample elements is chosen; then its distribution is determined under the null hypothesis. Now given a probability $1 - \alpha$ close to 1 (α small), we determine an interval (T_1, T_2) in such a way that under the validity of the simple hypothesis H_0 the value of T will belong to this interval with probability $1 - \alpha$:

$$P(T_1 < T < T_2 \mid H_0) = 1 - \alpha,$$

while it will fall outside the interval with the small probability

$$P(T \leqslant T_1 \text{ or } T \geqslant T_2 \mid H_0) = \alpha.$$

If the value of T computed from the observed sample is in (T_1, T_2) then H_0 will be accepted and it will be rejected, on the level $1 - \alpha$, otherwise. In the first case we say: *the observed data do not contradict the validity of* H_0.

In certain cases, depending on the practical situation, the hypothesis is *one-sided*. For example, in (a) it has the form

$$H_0\colon\ p \leqslant p_0 = 0.02,$$

$$H_1\colon\ p > p_0 = 0.02.$$

Accordingly, the acceptance region and critical region will show a similar shape; if we denote by $\varkappa_n$ the number of defective items in a sample of size n taken from the lot, then we shall obtain for the critical region:

$$\mathfrak{X}_r = \{\varkappa_n \geqslant c\},$$

while for the acceptance region

$$\mathfrak{X}_a = \{\varkappa_n < c\},$$

where c satisfies the relation

$$P(\varkappa_n < c \mid H_0) = 1 - \alpha.$$

4.3.2 Errors of the First and Second Kind

As we have already seen, a statistical decision is always based on a sample; accordingly the test procedure may lead to wrong decisions due to the, random character of the test statistic computed from the observed sample elements. We say that an *error of the first kind* is committed if we reject the null hypothesis although it is true. An *error of the second kind* occurs if we accept H_0 in spite of the fact that the alternative hypothesis is valid. We may display the different possible outcomes of a decision in a way shown in Table 6a.

TABLE 6a

	The null hypothesis H_0 will be	
	accepted	rejected
H_0 is true	good decision	wrong decision (error of first kind)
H_1 is true	wrong decision (error of second kind)	good decision

From the practical point of view the main question is how often we make mistakes in repeated applications of the test procedure. For this reason we have to determine the probabilities of the errors of first and second kind.

Turning to the example treated in the preceding section, the probability of an error of the first kind is

$$P(\mathfrak{X}_c \mid H_0) = P(u < -2.58 \text{ or } u > 2.58) = 0.01.$$

This follows from the definition: we make an error of first kind if H_0 is true but the sample results are in the critical region. This is expressed in our relation. This means that the probability of an error of the first kind in our test is identical to what we called the size of the test.

On the other hand, an error of second kind is made if the alternative hypothesis H_1 is true but we accept H_0. In our case the hypothesis H_1 is a composite one, containing a set of distributions instead of one unique distribution. Consequently, the probability of an error of the second kind can be determined only if the value of the parameter μ, i.e. the true distribution, is specified. Accordingly, the probability of the error of the second kind is a function of the parameter values belonging to H_1. Let us denote this function for the example in question by $p_{\mathrm{II}}(\mu)$; this can be expressed according to the definition in the following way:

$$p_{\mathrm{II}}(\mu) = P(-\,2.58 < u < 2.58 \mid \mu).$$

We note that when the true value of the parameter is $\mu (\neq \mu_0)$, the statistic

$$u = \frac{\overline{X} - \mu_0}{\sigma_0} \sqrt{n}$$

is no longer distributed according to $N(0.1)$. However, rearranging the form of u in the following way,

$$u = \frac{\overline{X} - \mu}{\sigma_0}\sqrt{n} + \frac{\mu - \mu_0}{\sigma_0}\sqrt{n} = \frac{\overline{X} - \mu}{\sigma_0}\sqrt{n} + \Delta, \quad \text{with} \quad \Delta = \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\,,$$

we observe that u is distributed according to $N(\Delta,\ 1)$. Thus $u - \Delta$ has the distribution $N(0.1)$. Hence, we obtain for the probability of an error of the second kind

$$p_{\mathrm{II}}(\mu) = P\left(-\,2.58 - \Delta < \frac{\overline{X} - \mu}{\sigma_0}\sqrt{n} < 2.58 - \Delta \mid \mu\right) =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-2.58-\Delta}^{2.58-\Delta} e^{-\frac{t^2}{2}}\, dt = \Phi(2.58 - \Delta) - \Phi(-\,2.58 - \Delta)\,.$$

If Δ is small, i.e. near to 0, the value of $p_{\mathrm{II}}(\mu)$ lies close to $1-0.01 = 0.99$. This means that if μ is close to μ_0 then the probability of the error of second kind turns out to be rather high. This is quite obvious: it is difficult to make the difference between parameter values lying near each other. In connection with this fact we shall make some remarks.

(a) The first comment concerns the null hypothesis and the probability of the error of the first kind. In practical applications it rarely occurs that the null hypothesis refers to a single parameter value; it is usually composite. E.g., a certain tolerance is given for the length of the cylinder, consequently, the null hypothesis has the form:

$$H_0'\colon\ \mu_0 - \delta_1 \leqslant \mu \leqslant \mu_0 + \delta_2,\ (\delta_i > 0,\ i = 1, 2).$$

In this case the probability of error of the first kind depends on the true value of μ if it belongs to H_0'. We now say that the test is of size α ($= 0.01$), if this probability never exceeds the value α, but attains it for at least one specified value of the parameter:

$$P(\mathfrak{X}_c \mid \mu) \leqslant \alpha, \quad \text{for all } \mu \text{ in } H_0'.$$

(b) In the case of the composite null hypothesis H_0' the alternative hypothesis has the form

$$H_1'\colon \mu < \mu_0 - \delta_1 \text{ or } \mu > \mu_0 + \delta_2.$$

For values of μ belonging to H_1', but near to $\mu_0 - \delta_1$ or $\mu_0 + \delta_2$, the probability of an error of the second kind will be high. It is common practice in such cases to alter the range of parameter values belonging to the alternative hypothesis, choosing $\delta_1'(> \delta_1)$ and $\delta_2'(> \delta_2)$ and replacing H_1' by

$$H_1''\colon \mu \leqslant \mu_0 - \delta_1' \text{ or } \mu \geqslant \mu_0 + \delta_2'.$$

Values of μ smaller than or equal to $\mu_0 - \delta_1'$ are considered to be very unfavorable, and the same holds for values larger than or equal to $\mu_0 + \delta_2'$. For these values of μ it is desirable to reject the alternative H_0' with high probability. How to construct the appropriate test will be indicated next.

(c) As we have seen, the probability of the error of second kind for given $\mu \in H_1''$ can be expressed in the following way:

$$p_{\text{II}}(\mu) = \int_{-2.58-\varDelta}^{2.58-\varDelta} e^{-\frac{t^3}{2}}\,dt;\ \mu < \mu_0 - \delta_1' \text{ or } \mu > \mu_0 + \delta_2'$$

where $\varDelta = \dfrac{\mu - \mu_0}{\sigma_0}\sqrt{n}$; this means that this probability depends on μ through $\varDelta$.

The values of $p_{\text{II}}(\mu)$ are illustrated in Fig. 24; a little analytical consideration shows that this value decreases if $\varDelta$ moves from zero in either direction. For given μ the value of $\varDelta$ can be increased in two different ways. One is purely professional-technical: diminishing the standard deviation by means of improving the production process. The other possibility is statistical: increase in the sample size n.

We turn now to the question of the *choice of error probabilities.*

According to the construction of the test procedure, we first fix the probability of error of the first kind by checking the level of the test. After this, the critical region will be assigned with the aid of a suitably chosen test statistic and sample size. From these, for each given value of μ belonging to the alternative, the probability of the error of second kind is also determined. This means that the probability of the error of second kind depends on the value of the probability of the error of first kind. A decrease in the probability of error of the first kind involves an increase in the probability of the error of second kind. When the critical region is decreased its complementary, i.e. the acceptance region, will be increased; but for a given distribution a larger probability belongs to a larger region.

This consideration is true for any fixed value of μ ($\in H_1'$). Figure 25 shows that the probability of error of the second kind is maximum when $\mu = \mu_0 - \delta_1'$ or $\mu = \mu_0 + \delta_2'$ (depending on which of these values is closer to μ_0).

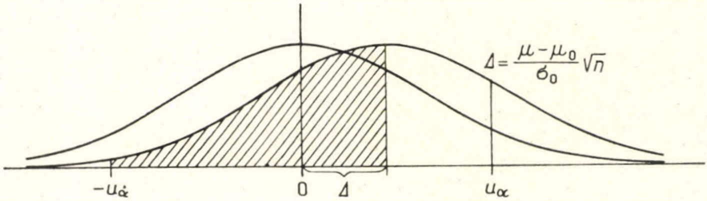


FIG. 24. The probability of an error of second kind (shaded area) in the case of the u test for a given alternative

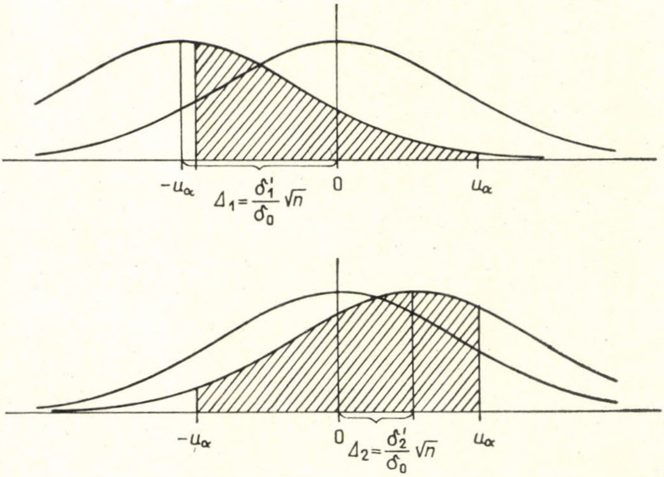


FIG. 25. Probabilities of the errors of second kind (shaded areas) in the case of the two-sided u test for the two extreme alternatives

From the previous considerations it follows that, in determining the error probabilities, we have to take into account the professional aspects as well: the practical consequences of errors of the first kind and of the second kind.

In applying statistical quality control methods, two extreme cases often occur: one kind of error involves much graver consequences than the other. Because of this, the probability of the former will be chosen as small as possible. If, e.g., a defective item produced by a machine will seriously damage the instrument in which it is mounted, or if a medication not fitting the prescription may damage one's health, then the probability of error of the second kind has to be set as low as possible.

In such cases the probability of error of the first kind will be higher: in the first example the production process will be interrupted more often than is necessary, while in the second example, good lots will be rejected with a relatively high frequency. On the other hand, in the majority of practical cases a slight insufficiency in the quality has no serious consequences, whereas the rejection of a good lot is very disadvantageous. In such cases, then, the probability of the error of first kind will be kept on a low level.

In other cases economic or financial considerations may lead to a suitable solution of the problem: in this respect the method given in the theory of decision functions (see Section 4.5) proves useful.

4.3.3 The Power Function and the OC Curve

We may get a good survey of a statistical test with the aid of its *power function*, which we will now define. In the field of statistical quality control and especially in the case of acceptance plans the so-called *operating characteristic function* is preferred over the power function. Both will be introduced and illustrated with example (b) and using the composite null hypothesis:

H_0: X is distributed according to $N(\mu, \sigma_0)$ with σ_0 given, and $\mu_0 - \delta \leqslant \mu \leqslant \mu_0 + \delta$ against the composite alternative

H_1: X is distributed according to $N(\mu, \sigma_0)$ with σ_0 given as above and $\mu < \mu_0 - \delta$ or $\mu > \mu_0 + \delta$.

First, we construct the critical region.

Given the level $1 - \alpha$, the test will be based on the mean value of the sample elements:

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

or more precisely on

$$u = \frac{\bar{X} - \mu_0}{\sigma_0}\sqrt{n}\,.$$

The critical region will be chosen symmetrically around μ_0, with $-u_\alpha$ and u_α:

$$\mathfrak{X}_c = \left\{\overline{X} : \overline{X} < \mu_0 - u_\alpha \frac{\sigma_0}{\sqrt{n}} \text{ or } \overline{X} > \mu_0 + u_\alpha \frac{\sigma_0}{\sqrt{n}}\right\}$$
$$= \{u : u < -u_\alpha \text{ or } u > u_\alpha\}.$$

To satisfy the requirement concerning the level of the test the following relation must hold:

$$P(\mathfrak{X}_c \mid \mu) \leqslant \alpha \text{ for } \mu_0 - \delta \leqslant \mu \leqslant \mu_0 + \delta. \tag{22}$$

As can be seen, this will be satisfied if the equality

$$P(\mathfrak{X}_c \mid \mu_0 - \delta) = P(\mathfrak{X}_c \mid \mu_0 + \delta) = 1 - \alpha$$

holds, or in another form, if

$$1 - P(\mathfrak{X}_c \mid \mu_0 \pm \delta) = 1 - P(-u_\alpha < u < u_\alpha \mid \mu_0 \pm \delta) = \alpha.$$

Under the assumption that the true value of the expectation is $\mu_0 - \delta$, the random variable

$$u = \frac{\overline{X} - \mu_0}{\sigma_0}\sqrt{n} = \frac{\overline{X} - (\mu_0 - \delta)}{\sigma_0}\sqrt{n} - \frac{\delta}{\sigma_0}\sqrt{n}$$

is distributed according to $N\left(-\frac{\delta}{\sigma_0}\sqrt{n}\,,\ \sigma_0\right)$, consequently, the following relation must hold:

$$P(\mathfrak{X}_c \mid \mu_0 - \Delta) =$$
$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\mu_\alpha - \frac{\delta}{\sigma_0}\sqrt{n}}^{u_\alpha - \frac{\delta}{\sigma_0}\sqrt{n}} e^{-\frac{t^2}{2}}\, dt = \Phi\left(u_\alpha - \frac{\delta}{\sigma_0}\sqrt{n}\right) - \Phi\left(-u_\alpha - \frac{\delta}{\sigma_0}\sqrt{n}\right).$$

As δ, σ_0 and n are given quantities, the unknown u_α can be determined by using Table IV (p. 385). As a consequence of the symmetry of the normal frequency curve, the relation $P(\mathfrak{X}_c \mid \mu_c + \delta)$ will be satisfied as well.

Having determined the critical region for the level $1 - \alpha$, let us consider the probabilities of the errors.

We begin with the consideration of the following function of μ

$$E_n(\mu) = P(\mathfrak{X}_c \mid \mu), \tag{23}$$

i.e., the probability of the critical region, i.e., the probability of the rejection of H_0, if the true value of the parameter is μ.

The function (23) is called the *power function* of the test; it is defined for each possible value of the parameter (Fig. 26).

If $\mu \in H_0$, then $E_n(\mu)$ is the probability of the error of first kind if the true value of the parameter is μ. According to Eq. (22), it holds that $E_n(\mu) \leqq \alpha$ for $\mu \in H_0$ as shown in the figure.

If μ is element of the alternative hypothesis, then $E_n(\mu)$ is the probability of a good decision: rejection of H_0 if H_0 is not true. The larger this value, the better the test for this μ; it is called the *power of the test* against the alternative μ.

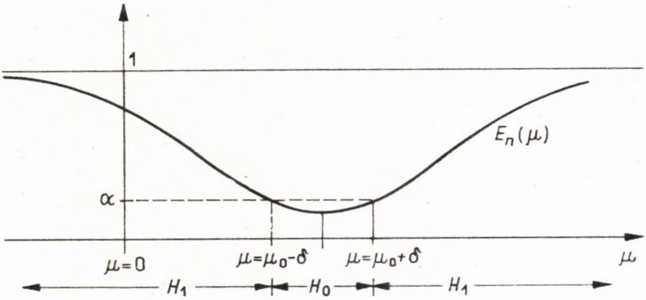


FIG. 26. Power function of the u test

Obviously, the probability of the error of second kind will be obtained in the following way:

$$p_{\text{II}}(\mu) = 1 - E_n(\mu), \quad \mu \in H_1.$$

While the power function gives the probability of the rejection of H_0, the *OC* (*operating characteristic*) *function* is defined as the probability of the acceptance of H_0 as a function of μ; it has the form

$$L(\mu) = P(\mathfrak{X}_a | \mu) = 1 - P(\mathfrak{X}_c | \mu) = 1 - E_n(\mu)$$

defined for each value of μ. It plays an important role in the construction and application of acceptance plans (see Chapter III, Section 2).

4.3.4 Uniformly Best Test. Lemma of Neyman and Pearson

As we have seen, a test is determined by the choice of the critical region and is performed by means of a test statistic. The test statistic, however, does not uniquely determine the test. For example, in the first case we choose a symmetric acceptance region $(-u_\alpha, u_\alpha)$ for the statistic

u, but we could also have chosen some nonsymmetric region with the property

$$(P - u_{\alpha 1} < u < u_{\alpha 2} \mid \mu_0) = \alpha.$$

The critical region $\{u : u < -u_{\alpha 1} \text{ or } u > u_{\alpha 2}\}$ provides a test procedure for the same problem on the same level. Furthermore, instead of the statistic

$$u = \frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n},$$

it is possible to select a different test statistic, based for example on the median of the sample [see Section 3.2.2.3]:

$$m_e = \begin{cases} X^*_{m+1}, & \text{if} \quad n = 2m + 1. \\ \dfrac{X^*_m + X^*_{m+1}}{2}, & \text{if} \quad n = 2m, \end{cases}$$

which is also an unbiased estimator of the expected value.

For an appropriate selection of the test to be used we must be able to compare two test procedures. The power function is a very convenient means for this purpose.

Let us consider again the problem of the hypotheses

$$H_0\colon \mu_0 - \delta \leqslant \mu \leqslant \mu_0 + \delta$$

$$H_1\colon \mu < \mu_0 - \delta \text{ or } \mu > \mu_0 + \delta$$

for the expected value μ of the length X of the cylinder produced by a given machine. In both H_0 and H_1, X is normally distributed with dispersion σ_0.

We have constructed the u test for making a decision between H_0 and H_1. Let us denote this test for the level $1 - \alpha$ by U_α, and its power function by

$$E_n(\mu;\ U_\alpha).$$

As U_α is a test with level $1 - \alpha$, the power function satisfies the relation

$$E_n(\mu;\ U_\alpha) \leqslant \alpha, \text{ for } \mu_0 - \delta \leqslant \mu \leqslant \mu_0 + \delta.$$

Let us suppose that we have constructed a test for the problem H_0 against H_1 based on the statistic m_e. Denote the test for level $1 - \alpha$ by M_α, and its power function by $E_n(\mu;\ M_\alpha)$. This test also has the property

$$E_n(\mu;\ M_\alpha) \leqslant \alpha, \text{ for } \mu_0 - \delta \leqslant \mu \leqslant \mu_0 + \delta.$$

The test U_α is considered better than the test M_α against a given alternative μ if the power of U_α is larger than that of M_α for this value μ. It turns out in this special case that this is true for each value of $\mu \in H_1$, i.e.

$$E_n(\mu;\ U_\alpha) \leqslant E_n(\mu;\ M_\alpha) \text{ for } \mu < \mu_0 - \delta$$
$$\text{or } \mu > \mu_0 + \delta.$$

In this case we say that the test U_α is *uniformly better* than the test M_α.

In general: having two tests T_1 and T_2 for deciding between the hypotheses H_0 and H_1, we say that T_1 is uniformly better than T_2 on the level $1 - \alpha$, if the following statements are satisfied: (a) the power functions of the tests T_1 and T_2 do not exceed the value α if the parameter belongs to H_0; (b) the power of T_1 is not less than the power of T_2 for each value of the parameter belonging to H_1; (c) there exists at least one value of the parameter belonging to H_1 for which the power of T_1 exceeds the power of T_2.

Returning to the example, we emphasize the following two circumstances: (1) Taking, for example a sample of size $n = 11$, it is much easier to determine the sample median m_e than to compute the value of $\bar{X}$. This means that, when using the test frequently, we sometimes prefer the use of the median (depending naturally on the consequences of wrong decisions). (2) The statistic $\bar{X}$ is a best linear unbiased estimator for the expected value and, in cases of normal, exponential and Poisson distribution, it is even a best unbiased estimator as a sufficient statistic.

We have stated (without proof) that the U_α test is uniformly better than the M_α test for the given problem. It is noteworthy that for making a decision between H_0 and H_1 the corresponding u-test is uniformly better than any other $1 - \alpha$ level test; it is a *uniformly most powerful test*. (Figure 27a.)

For a given decision problem there does not always exist a uniformly most powerful test. We often have situations like that shown in Fig. 27b: for certain values of the parameter belonging to the alternative one of the tests is more powerful; while for other values the other test is more powerful.

There is, however, an (exceptional) case when a uniformly best test always exists: namely when both H_0 and H_1 are simple hypotheses, i.e., both consist of a single distribution. The corresponding statement and construction of the uniformly best test statistic is contained in the following lemma due to Neyman and Pearson. We formulate this for continuous random variables, merely remarking on the discrete case.

Lemma of Neyman and Pearson: For the distribution function of the random variable X assume that we have the null hypothesis

$$H_0\colon\ P(X \leqslant x) = F(x) = \int_{-\infty}^{x} f(t)\,dt$$

and the alternative

$$H_1\colon\ P(X \leqslant x) = G(x) = \int_{-\infty}^{x} g(t)\,dt.$$

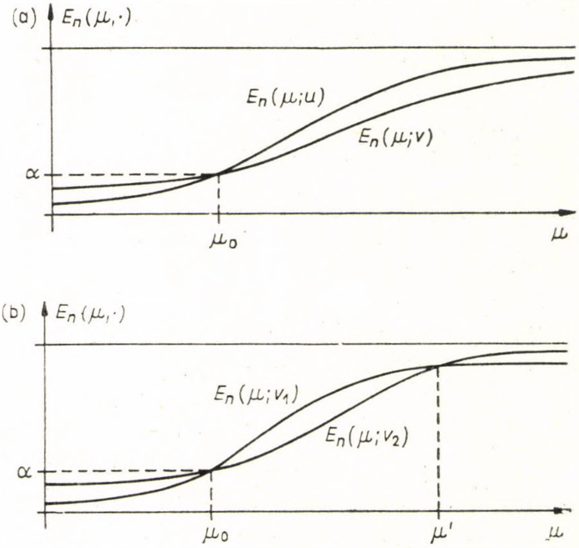


FIG. 27. Comparison of power functions belonging to one-sided tests; (a) the u test is uniformly better than the v test, (b) test v_1 is better in (μ_0, μ') test, v_2 is better for $\mu > \mu'$

For making a decision between H_0 and H_1 on the level $1 - \alpha$ based on the sample $X_1, X_2, \ldots, X_n$, the uniformly best test is obtained by means of the test statistic

$$t(X_1\, X_2, \ldots\, X_n) = \frac{g(X_1)\, g(X_2) \ldots g(X_n)}{f(X_1)\ f(X_2) \ldots f(X_n)}$$

with the critical region

$$\mathfrak{X}_c = \{(X_1, X_2, \ldots X_n)\colon\ t(X_1, X_2, \ldots ;\ X_n) < c_\alpha\}$$

where the constant c_α is determined by the following relation:

$$P(\mathfrak{X}_c \mid H_0) = P(t(X_1, X_2, \dots X_n) \geqslant c_\alpha \mid H_0) = \alpha.$$

This is called the *probability ratio test.*

Without giving the proof we turn now to the following important example which is closely connected with our earlier considerations.

Let $F(x)$ and $G(x)$ be normal distribution functions with the common variance σ_0 and with the expectations μ_0 and $\mu_1 (\mu_0 < \mu_1)$, respectively:

$$H_0\colon\ P(X \leqslant x) = F(x) = \Phi(x;\ \mu_0,\ \sigma_0),$$

$$H_1\colon\ P(X \leqslant x) = G(x) = \Phi(x;\ \mu_1,\ \sigma_0).$$

Then

$$t(X_1, X_2, \dots X_n) = \frac{\left(\frac{1}{\sqrt{2\pi}\,\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_1^n (X_i-\mu_1)^2}}{\left(\frac{1}{\sqrt{2\pi}\,\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_1^n (X_i-\mu_0)^2}} = e^{\frac{n}{\sigma_0^2}(\mu_1-\mu_0)\bar{X} - \frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)}.$$

For the critical region we obtain

$$\mathfrak{X}_c = \left\{\bar{X} : \frac{n}{\sigma_0^2}(\mu_1 - \mu_0)\bar{X} - \frac{n}{2\sigma_0^2}(\mu_1^2 - \mu_0^2) \geqslant \log c_\alpha \right\},$$

which can be reduced to the form

$$\mathfrak{X}_c = \{\bar{X} > c'_\alpha\} = \left\{u = \frac{\bar{X} - \mu_0}{\sigma_0}\sqrt{n} > \frac{c'_\alpha - \mu_0}{\sigma_0}\sqrt{n} = c''_\alpha\right\},$$

where

$$c''_\alpha = c''_\alpha(n,\ \mu_0,\ \mu_1,\ \sigma_0).$$

Thus we obtained the u statistic, and when H_0 is true this is distributed according to $N(0,1)$. Having $1 - \alpha$ level test, the relation

$$P(\mathfrak{X}_c \mid H_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma_0}\sqrt{n} > c''_\alpha \mid \mu_0\right) = \alpha$$

must hold. This means that *the value of c''_α is determined by the relation*

$$1 - \Phi(c''_\alpha) = \alpha;$$

consequently, it does not depend on μ_1 or on the other constants, but on α only. This fact has the significant consequence that this *test is uniformly most powerful,* not only for deciding between H_0 and H_1, but for deciding *between the simple null hypothesis*

$$H_0\colon\ P(X \leqslant x) = \Phi(x;\ \mu_0,\ \sigma_0)$$

and the composite alternative

$$H_1^*\colon P(X \leqslant x) = \Phi(x;\ \mu,\ \sigma_0),\ \mu > \mu_0$$

as well.

Similar considerations show that the most powerful property remains valid for the composite null hypothesis

$$H_0^*\colon P(X \leqslant x) = \Phi(x;\ \mu,\ \sigma_0),\ \mu \leqslant \mu_0,$$

against the alternative

$$H_1^*\colon P(X \leqslant x) = \Phi(x;\ \mu,\ \sigma_0),\ \mu < \mu_0.$$

Here we have an example of a uniformly most powerful test for composite hypotheses.

Finally, we wish to remark that in the discrete case it is not always possible to construct a test with a given level $1 - \alpha$. For practical purposes it suffices, however, to approximate α from below or from above, according to the practical situation. Another possibility is the use of so-called randomization; this is discussed in the following section.

4.3.4.1 On Randomized Tests. Lots of products are examined and requirements concerning their fraction defective are set, in the following way:

$$H_0\colon p \leqslant p_0 = 0.04,$$

$$H_1\colon p > p_0 = 0.04.$$

Our aim is to accept lots with no more than 4% fraction defective, with probability at least $1 - \alpha = 95\%$. Taking into account the difficulties in taking the sample elements from the lot and the costs of examining the single pieces, a sample of size $n = 25$ is advisable.

In constructing a good test we start with the simple null hypothesis

$$H_0\colon p = p_0 = 0.04$$

and the simple alternative

$$H_1'\colon p = p_1 (> p_0 = 0.04),$$

where p_1 is any fixed fraction larger than p_0. Applying the Neyman–Pearson lemma, the test statistic

$$t(k) = \frac{\binom{n}{k} p_1^k (1 - p_1)^{n-k}}{\binom{n}{k} p_0^k (1 - p_0)^{n-k}} \tag{24}$$

will be used, where k refers to the number of defective items in the sample of size $n = 25$. The critical region will have the form $\{t(k) \geqslant C\}$, for which the relation

$$P(t(k) > C \mid H_0) = \alpha = 0.05 \tag{25}$$

must hold. From (24), the condition $t\,(k) \geqslant C$ leads to the form

$$k \geqslant \frac{\log C - n \log \dfrac{1-p_1}{1-p_0}}{\log \dfrac{p_1}{p_0} \dfrac{1-p_0}{1-p_1}} = C_1(C, n, p_0, p_1) = C_1 .$$

According to (24), the constant C_1 satisfies the relation

$$P(k \geqslant C_1 \mid p_0 = 0.04) = \alpha = 0.05. \tag{26}$$

This means that C_1 and the critical region do not depend on the value of $p_1\ (> p_0)$, and we have again a uniformly best test between the composite alternatives

$$H_0^*\colon\ p \leqslant p_0 = 0.04$$

and

$$H_1^*\colon\ p > p_0 = 0.04.$$

Unfortunately, we cannot satisfy relation (26) since we find in the Table VIb (p. 391) of the binomial distribution for $n = 25$, $p = 0.04$:

$$P(k \geqslant 3 \mid p_0 = 0.04) = 0.0765$$

$$P(k \geqslant 4)\ p_0 = 0.04) = 0.0165.$$

This means that, in this case, only levels 92.35% or 98.35% can be realized.

If we insist on the 95% level, we may turn to the following so-called randomization:

The critical region will be constructed in such a way that

we reject H_0 certainly, if $k > 3$,

we reject H_0 with probability q $(0 < q < 1)$, if $k = 3$,

we accept H_0 certainly, if $k < 3$,

where the value of q is to be determined so that the probability of rejecting of H_0, if it is true, should have the value $\alpha = 0.05$. (How to proceed in the second case ($k = 3$) will be treated later.)

For the determination of the probability, we formulate the test procedure with the aid of the so-called *test function* $\Phi(k)$. *$\Phi(k)$ is the probability of the rejection of* H_0, the value of which depends on k and has the form

$$\Phi(k) = \begin{cases} 1, & \text{if} \quad k > 3, \\ q, & \text{if} \quad k = 3, \\ 0, & \text{if} \quad k < 3. \end{cases} \tag{27}$$

The probability of rejecting the lot will depend as well, of course, on the true value p of the fraction defective. Using the rule of total probabilities (Section 1.9) and taking into account that $\Phi(k)$ is a conditional probability:

$\Phi(k) = P$ (rejection of H_0 when the number of the defective items in the sample is k); we have

P (rejection of H_0 when p_0 is the fraction defective) $=$

$$\sum_{k=0}^{n} \Phi(k)\, P(k \mid p) =$$

$$1 \cdot P(k > 3 \mid p) + qP(k = 3 \mid p) + 0 \cdot P(k < 3 \mid p) =$$

$$P(k > 3 \mid p) + qP(k = 3 \mid p).$$

This must be equal to 0.05 if $p = p_0 = 0.04$:

$$P(k > 3 \mid p_0 = 0.04) + qP(k = 3 \mid p_0 = 0.04) = 0.05$$

or $0.0165 + 0.0500q = 0.05$.

From this we obtain the value $q = 0.558 = 55.8\%$.

This means that if the sample contains $k = 3$ defective items we will reject H_0 with probability 55.8% by means of the following random experiment: We put into a box 558 red balls and 442 white balls (or, approximating these ratios, 56 red balls and 44 white ones). Then we randomly draw one ball out of the box. If this is red then we reject H_0, if it is white then we accept H_0.

We formulate our randomization procedure in a slightly more general form:

We have to decide at level $1 - \alpha$ between the hypotheses

$$H_0\colon \vartheta = \vartheta_0, \text{ and}$$

$$H_1\colon \vartheta = \vartheta_1.$$

Let the random variable X have density function $f(x;\ \vartheta)$ with $\vartheta = \vartheta_0$ or ϑ_1. Then the Neyman–Pearson lemma leads to the test function $t(X) = \dfrac{f(X;\vartheta_1)}{f(X;\vartheta_0)}$.

In certain cases it turns out that, for a constant C, the relation

$$P(t(X) > C \mid H_0) < \alpha < P(t(X) \geqslant C \mid H_0)$$

holds, which means that for an exact level $1 - \alpha$ a critical region cannot be constructed.

In such cases we introduce a test function $\Phi(X)$ expressing the probability of the rejection of H_0 if the sample element X is observed; this will have the form:

$$\Phi(X) = \begin{cases} 1, & \text{if} \quad t(X) > C, \\ q, & \text{if} \quad t(X) = C, \\ 0, & \text{if} \quad t(X) < C, \end{cases}$$

where C and q are uniquely determined by the relation

$$P(\text{rejection of } H_0 | H_0 \text{ is true}) = 1P(t(X) > C | H_0) + q(t(X) = C | H_0) = \alpha.$$

If the observations result in $t(X) = C$, then the decision is carried out by means of a random experiment which yields the prescribed probability q for the rejection of H_0. This experiment has to be independent of the original phenomenon.

In practice minor change in the value of the test size would seem to be preferable over using randomized test. In such circumstances, for example, one of the following principles may be adopted: (1) The prescribed size of the test is considered as the upper bound for the actual test size. (2) The solution leading to smaller change in the test size is to be preferred.

In the above example the first principle leads to the critical region characterized by the inequality $k \geqslant 4$. According to the second principle, the critical region will be $k \geqslant 3$.

4.3.5 Consistency and Unbiasedness

In order to define and illustrate these two requirements in connection with statistical tests, we turn again to our previous example. For testing the hypothesis H_0: $\mu = \mu_0$ against the alternative H_1: $\mu \neq \mu_0$, the test was constructed with the critical region

$$\mathfrak{X}_c = \{u \leqslant -u_\alpha \text{ or } u \geqslant u_\alpha\}$$

on level $1 - \alpha$, i.e.

$$P(\mathfrak{X}_c \mid \mu = \mu_0) = \alpha.$$

In Section 4.3.2 the error of the second kind was determined, which leads to the power function

$$\begin{aligned} E_n(\mu) &= 1 - p_{\text{II}}(\mu) = 1 - \Phi(u_\alpha - \varDelta) + \Phi(-u_\alpha - \varDelta) \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int\limits_{-u_\alpha - \varDelta}^{\mu_\alpha - \varDelta} e^{-\frac{t^2}{2}} dt \end{aligned}$$

with

$$\varDelta = \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\,.$$

For a given value $\mu \neq \mu_0$, with an increasing sample size n, the value of $\varDelta$ increases; but as can be easily seen (e.g., by differentiating $E_n(\mu)$ with respect to $\varDelta$) the power $E_n(\mu)$ is an increasing function of n:

$$E_n(\mu) \leqslant E_{n+1}(\mu).$$

Furthermore, if $n \to \infty$, then $\varDelta \to \infty$ and the following relation is valid:

$$\lim_{n \to \infty} E_n(\mu) = 1 \quad \text{for fixed} \quad \mu (\neq \mu_0).$$

This means that for each element of the composite alternative, the power of the test may be chosen arbitrarily close to one if only the sample size can be made large enough, i.e., the error of second kind can be diminished as far as we want by increasing the sample. The same is true for the test between the hypotheses

$$H_0 \mu_0 - \delta \leqslant \mu \leqslant \mu_0 + \delta \text{ and } H_1\colon \mu < \mu_0 - \delta \text{ or } \mu > \mu_0 + \delta.$$

When speaking about a test, we usually think of a test which is defined for a sequence of sample sizes: $n_0, n_0 + 1, \ldots$ The situation is somewhat different in the case of the sequential procedure (see Chapter III; Section 2.2.3). If a test has the property that for each fixed alternative the power of the test tends to 1 as n tends to infinity, then *the test* is called *consistent*.

It is obvious that if μ lies near μ_0 and only a small error of second kind is allowed, then the required sample size n will be very large.

A test of level $1 - \alpha$ is *unbiased* if the following is true: for each element of the alternative the test rejects H_0 with probability greater than α. In other words: the test rejects H_0 with larger probability when H_0 is not true than it does when it is valid. Expressed in terms of the power function in this case

$$E_n(\mu) \geqslant \alpha, \text{ if } \mu \in H_1.$$

As Fig. 26 shows, this test is unbiased. Nonparametric tests with a wide set of alternatives show biasedness for certain distributions belonging to the alternative.

4.3.6 Parametric Tests of Various Hypotheses Involving Gaussian Variables

4.3.6.1 General Remarks. Let X be distributed according to the normal law:

$$P(X \leqslant x) = \Phi(x;\ \mu, \sigma).$$

In the sections which follow, several tests of different hypotheses on μ and σ are described. Both one- and two-sided tests are introduced and examples of appropriate applications are given. Some cases illustrate the comparison of the expected values or variances of two or more Gaussian variables $X^{(1)}, X^{(2)}, \ldots, X^{(r)}$, $r \geqslant 2$. In each case the null-hypothesis is specified and the test is given in terms of the critical region determined by the selected alternatives. The name of a test usually agrees with the name of the test statistic used in it.

All the tests presented are consistent and unbiased. Some are uniformly most powerful among all possible unbiased tests. In certain cases the test will be uniformly most powerful. Concerning these important questions we refer to the book by Lehmann [A 30].

Strictly speaking, these tests are appropriate only if the underlying variables have a Gaussian distribution. When normality is doubtful, a goodness-of-fit test for normality is necessary; these will be described in Section 4.3.7.3. We note, however, that small departures from normality do not cause significant changes in the probabilities of errors of the first and second kind.

For practical use we include some of the most frequently needed tables. References to more specialized tables are also given. With their aid, the critical regions of the tests can be determined in terms of the so-called critical values of the test statistics for several levels of the test. To obtain probabilities of the errors of the second kind for given alternatives further tables are needed in most cases; we shall sometimes refer to them as well.

4.3.6.2 The *u* Test. *The one-sample problem:* We have already treated this test in the preceding paragraphs, when the concepts associated with a test procedure were introduced.

Now we summarize the main features of the test:

For the one-sample problem we shall assume throughout that X is distributed according to the normal law with known dispersion σ_0. The test will be based on the arithmetic mean $\overline{X}$ computed from n independent observations $X_1, X_2, \ldots, X_n$ of X. The level $1 - \alpha$ is to be prescribed in each case.

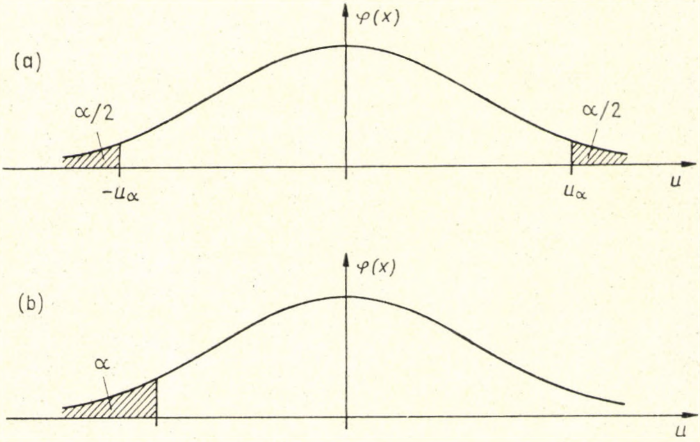


FIG. 28. $1 - \alpha$ level critical regions for the two-sided (a) and for the one-sided (b) *u* tests

For testing the null hypothesis

$$H_0\colon\ E(X) = \mu_0$$

the statistic

$$u = \frac{\overline{X} - \mu_0}{\sigma_0}\sqrt{n}$$

will be applied. If the null hypothesis is valid, u is distributed according to $N(0, 1)$ (Fig. 28).

Against the alternative

$$H_1\colon\ E(X) = \mu \neq \mu_0,$$

the critical region

$$\mathfrak{X}_c = \{u \leqslant -u_\alpha \text{ or } u \geqslant u_\alpha\}$$

is obtained; here the value of u_α can be determined from the relation

$$2\,\Phi(u_\alpha) - 1 = 1 - \alpha \text{ or } \Phi(u_\alpha) = 1 - \frac{\alpha}{2}$$

with the aid of the Gaussian probabilities in Table IV (p. 385).

Against the alternative

$$H_1'\colon\ E(X) = \mu > \mu_0,$$

the critical region

$$\mathfrak{X}_c = \{u \geqslant u(\alpha)\}$$

yields the test at level $1 - \alpha$, if for $u(\alpha)$ the following relation holds:

$$\Phi(u(\alpha)) = 1 - \alpha.$$

This critical region gives on the same level a uniformly best test between the composite hypotheses

$$H_0'\colon\ E(X) = \mu \leqslant \mu_0 \text{ and } H_1'\colon\ E(X) = \mu > \mu_0.$$

The power function is in the case H_1

$$E_n(\mu) = 1 - \Phi(u_\alpha - \Delta) + \Phi(-u_\alpha - \Delta),$$

while in the case of H_1'

$$E_n'(\mu) = 1 - \Phi(u(\alpha) - \Delta),$$

where for both cases $\Delta = (\mu - \mu_0)\sqrt{n}\,/\,\sigma_0$.

Consistency and unbiasedness hold in both cases.

The two-sample problem: Let the random variables $X^{(1)}$ and $X^{(2)}$ be distributed according to the normal law with the known dispersions σ_1 and σ_2, respectively. The null hypothesis

$$H_0\colon\ E(X^{(1)}) = E(X^{(2)})$$

is to be tested at the level $1 - \alpha$. The decision will be based on the samples $X_1^{(1)}, X_2^{(1)}, \ldots, X_{n_1}^{(1)}$ and $X_1^{(2)}, X_2^{(2)}, \ldots, X_{n_2}^{(2)}$, taken on $X^{(1)}$ and $X^{(2)}$, respectively. In denoting the corresponding sample means by $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$ the test statistic is given by

$$u = \frac{\overline{X}^{(1)} - \overline{X}^{(2)}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}};$$

in case H_0 is true, u is distributed according to $N(0, 1)$.

If the alternative is

$$H_1\colon\ E(X^{(1)}) \neq E(X^{(2)})$$

the critical region is given by

$$\mathfrak{X}_c = \{u \leqslant -u_\alpha \text{ or } u \geqslant u_\alpha\}$$

with

$$\Phi(u_\alpha) = 1 - \frac{\alpha}{2}\,.$$

For the alternative

$$H_1'\colon\ E(X^{(1)}) > E(X^{(2)})$$

the best critical region is given by

$$\mathfrak{X}_c' = \{u \leqslant -\,u(\alpha)\} \text{ with } \Phi(u(\alpha)) = 1 - \alpha.$$

The power function for the alternative $E(X^{(1)}) = E(X^{(2)}) + \delta$ in the case of H_1 is

$$E_n(\delta) = 1 - \Phi(u_\alpha - \Delta) + \Phi(-u_\alpha - \Delta).$$

In the case of H_1', when $\delta > 0$,

$$E_n'(\delta) = 1 - \Phi(u(\varepsilon) - \Delta).$$

In both cases

$$\Delta = \frac{\delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}\,.$$

Example: Cylinders are produced on a machine; by previous experience we know that the length of the cylinder is normally distributed with standard deviation $\sigma_0 = 0.012$ mm. At a certain time a sample of size 16 was taken, and, for the sample means of the lengths, the value 12.083 mm was obtained. One hour later a sample of size 9 was taken, and, for the arithmetic mean, the value 12.072 was found. A decision is to be made as to whether or not the expected value of the length has changed in the course of this hour. The level $1 - \alpha = 0.95$ will be used for the test.

Denoting by $X^{(1)}$ the length in the first sample we have, for the arithmetic mean of the sample of size $n = 16$, the value $\overline{X}^{(1)} = 12.083$. In the case of the second sample we denote the random variable by $X^{(2)}$ and the measuring of $m = 9$ elements resulted in the arithmetic mean $\overline{X}^{(2)} = 12.072$. For making a decision between the hypotheses

$$H_0\colon\ E(X^{(1)}) = E(X^{(2)})$$

and

$$H_1\colon\ E(X_{(1)}) \neq E(X^{(2)}),$$

we compute the value of the u statistic:

$$u = \frac{16.083 - 16.072}{0.012\sqrt{\frac{1}{16} + \frac{1}{9}}} = 2.2\,.$$

From $\Phi(u_{0.05}) = 1 - 0.025 = 0.975$, we obtain $u_{0.05} = 1.96$, i.e., the acceptance region will be $(-1.96;\ 1.96)$. As the observed value of u is inside the critical region, the null hypothesis, that the expected value of the length remained unchanged, will be rejected at level 95%.

4.3.6.3 The *F* Test. Let X and Y be normally distributed random variables. We want to know whether or not their variances are equal. The null hypothesis

$$H_0\colon\ D^{(2)}(X) = D^{(2)}(Y),$$

can be investigated with the aid of the *F test*. Taking from the two random variables the samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$, respectively, we determine the corresponding empirical variances s_n^2 and s_m^2. The test procedure differs slightly for the one-sided and two-sided cases and is as follows:

In the *one-sided case*, i.e., if the alternative hypothesis is

$$H_1^+\colon\ D^2(X) > D^2(Y),$$

we always reject H_1^+ (accept H_0) if $s_n^2 < s_m^2$ is valid for the observed values: the empirical variance of X being smaller than the empirical variance of Y, this cannot suggest otherwise. If, however, $s_n^2 > s_m^2$, we compute the statistic

$$F = \frac{s_n^2}{s_m^2} > 1\,.$$

Under the validity of H_0 the random variable $(n-1)F/(m-1)$ has a beta distribution of second kind with the pair of parameters $((n-1)/2, (m-1)/2)$ (see Section 2.9.9). The distribution of F is called F distribution with parameters (or degrees of freedom) $f_1 = n - 1$, $f_2 = m - 1$ (Fig. 29).

Now we determine the critical region for the level $1 - \alpha$.

$$\mathfrak{X}_c^* = \{F > F_{f_1,f_2;1-\alpha}\}.$$

Table VII (pp. 392–396) contains critical values for $\alpha = 0.05;\ 0.01;\ 0.001$

to which, in the one-sided case, there belong the levels 95%, 99% and 99.9%, respectively. If the observed value of F is in the critical region, we reject H_0; otherwise we say that the samples do not contradict the assumption of equal variance.

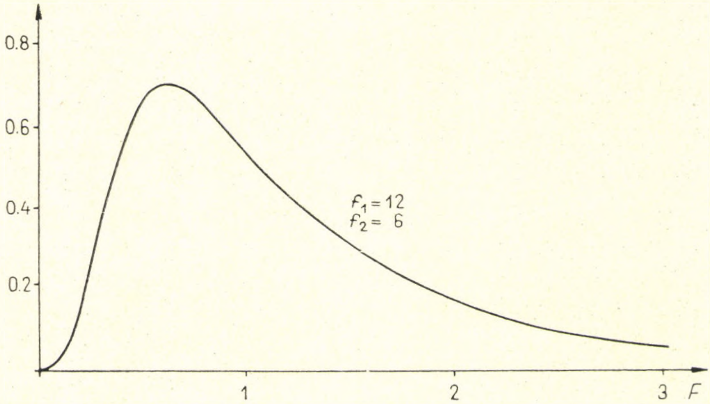


FIG. 29. Density function of the F distribution for $f_1 = 12$, $f_2 = 6$

In the *two-sided case*, i.e., if the alternative hypothesis is

$$H_1\colon\ D^2(X) \neq D^2(Y)$$

the $1 - 2\alpha$ level critical region of the test statistic

$$F = \frac{s_n^2}{s_m^2}$$

can be expressed, using the former notation, as

$$\mathfrak{X}_c = \{F > F_{f_1,f_2,1-\alpha}\} + \{1/F > F_{f_2,f_1,1-\alpha}\}.$$

Since the critical values are larger than 1, only the larger one of F and $1/F$ is to be computed; this means, we have to divide the larger variance by the smaller one.

In practice, usually the notations are chosen after the actual values of the variances are known and the sample with larger empirical variance will be considered to be the "first" i.e., we denote by F the ratio of the larger and smaller variances.

Example: In order to compare the working precision of two machines, the lengths of 80 cylinders from each machine were measured. If X

and Y denote the lengths of cylinders belonging to the two machines, the null hypothesis

$$H_0\colon D^2(X) = D^2(Y)$$

is to be tested against the two-sided alternative

$$H_1\colon D^2(X) \neq D^2(Y)$$

at the level $1 - 2\alpha = 0.90$.

For the values of the variances, $s_n^2 = 0.0968$ and $s_m^2 = 0.0805$ were obtained from the samples; in this case $n = m = 80$ and

$$F = \frac{0.0968}{0.0805} = 1.2015 .$$

In Table VII we find the critical value $F_{79.79;\,0.95} = 1.45$.

Since $F < F_{79.79;\,0.95}$, we accept the hypothesis of the identity of the two variances and consequently regard the two machines as working with the same accuracy.

4.3.6.4 The t-(Student)Test. The Welch Test. The u test given in Section 4.3.6.2 was appropriate if the variance(s) of the normal variable(s) was (were) known. However, knowing only that the random variable(s) considered is (are) from (a) Gaussian distribution(s), and having to make a decision concerning the expected values, the procedure is as follows:

The one-sample case: Let the random variable X have the distribution $N(\mu, \sigma^2)$ $(-\infty < \mu < \infty,\ \sigma > 0)$ and let us denote by $X_1, X_2, \ldots, X_n$ a sample of size n for X. For the null hypothesis

$$H_0\colon E(X) = \mu_0,$$

we construct the test statistic

$$t_{n-1} = \frac{\overline{X} - \mu_0}{s_n} \sqrt{n}$$

where X denotes the sample mean and s_n^2 the empirical variance. When H_0 is valid, t_{n-1} is distributed according to the Student distribution with parameter $n - 1$ (see Section 2.9.8).

For the alternative hypothesis

$$H_1^+\colon E(X) = \mu > \mu_0,$$

we choose the critical region of size α

$$\mathfrak{X}_c^+ = \{t_{n-1} > t(\alpha)\},$$

which is determined by the relation

$$P(t_{n-1} \leqslant t(\alpha) \mid H_0) = 1 - \alpha.$$

Critical values $t(\alpha)$ can be found in Table VIII (p. 397) at levels $(1 - \alpha) = 97.5\%$, 99.5%, 99.95%.

If the alternative hypothesis is two-sided, i.e.

$$H_1\colon\ E(X) = \mu \neq \mu_0,$$

then at the level $1 - \alpha$ the critical region will be

$$\mathfrak{X}_c = \{t_{n-1} < -t_\alpha \text{ or } t_{n-1} > t_\alpha\},$$

where for the value t_α the following relation holds:

$$P(-t_\alpha < t_{n-1} \leqslant t_\alpha \mid H_0) = 1 - \alpha.$$

Table VIII also gives the values of t_α for the levels $1 - \alpha = 95\%$, 99%, 99.9%.

The *power function of the Student test* has, in the two-sided case, the form

$$E_n(\mu;\, t_{n-1}) = 1 - P\left(-t_\alpha < \frac{\bar{X} - \mu_0}{s_n}\sqrt{n} \leqslant t_\alpha \,\middle|\, \mu\right).$$

Using again the notation σ^2 for the variance, χ_{n-1} for $\sqrt{n-1}\,\dfrac{s_n}{\sigma}$, Δ for the expression $\dfrac{\mu - \mu_0}{\sigma}\sqrt{n}$, and X' for the standardized variable $\dfrac{\bar{X} - \mu}{\sigma}\sqrt{n}$, the power function has the form

$$E_n(\mu;\, t_{n-1}) = 1 - P\left(-t_\alpha < \frac{X' + \Delta}{\chi_{n-1}}\sqrt{n-1} < t_\alpha \,\middle|\, \mu\right).$$

The first factor of the random variable $\dfrac{X' + \Delta}{\chi_{n-1}}\sqrt{u-1}$ is the quotient of a random variable distributed according to $N(\Delta, 1)$ and of a random variable having chi-distribution with parameter $n - 1$. The distribution of this expression is called *noncentral t distribution*. It is tabulated in [C 6]. The determination of the power function in the case of H_1^+ leads to a similar result.

The two-sample case: The t (Student) test can be used for the two-sample case *only if we may assume that the unknown variances of the two samples agree.* The verification of the validity of this assumption can be based on earlier experiences or, in the absence of such experience, the F test is to be applied.

Assuming the equality of the unknown variances, the procedure is as follows:

Let X and Y be random variables distributed according to the normal law and having equal variances: $D^2(X) = D^2(Y)$. The null hypothesis

$$H_0\colon\ E(X) = E(Y)$$

is to be investigated with the aid of the samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ taken on the random variables X and Y, respectively. In our usual notation the test statistic is the following:

$$t_{n+m-1} = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)\,s_n^2 + (m-1)\,s_m^2}} \sqrt{\frac{n\,m(n+m-2)}{n+m}}\,.$$

Under the validity of the null hypothesis this is distributed according to the t distribution with parameter $n + m - 2$.

Against the one-sided alternative

$$H_1^+\colon\ E(X) > E(Y)$$

at the level $1 - \alpha$, the critical region

$$\mathfrak{X}_c^+ = \{t_{n+m-2} > t(\alpha)\}$$

is to be applied. The critical value is determined by means of the relation

$$P(t_{n+m-2} \leqslant t(\alpha) \mid H_0) = 1 - \alpha.$$

Against the two-sided alternative

$$H_1\colon\ E(X) \neq E(Y)$$

at the level $1 - \alpha$, the critical region has the form:

$$\mathfrak{X}_c = \left\{t_{n+m-2} < -\,t_\alpha \text{ or } t_{n+m-2} > t_\alpha\right\}.$$

For the critical value the following relation holds:

$$P(-\,t_\alpha < t_{n+m-2} \leqslant t_\alpha \,|\, H_0) = 1 - \alpha\,.$$

The critical values in Table VIII mentioned in connection with the one-sample test are also used in the two-sample case. The power function leads in this case also to a noncentral t distribution.

As already pointed out in Section 2.9.8, the Student distribution converges to the Gaussian distribution as the parameter n tends to infinity. Therefore, for large samples, e.g., if the parameter of the t statistics exceed the value $n = 80$, the table of the Gaussian distribution can be used.

Example: In the preceding Section 4.3.6.3 the precision of two machines was compared. In that case we accepted the equality of the two variances of the cylinder lengths. Now we want to know whether the lengths (their expectations) are equal as well. We can now assume that the random variables X and Y are distributed according to the normal law with $D(X) = D(Y)$; the null hypothesis to be tested is

$$H_0: E(X) = E(Y)$$

and the two-sided alternative is

$$H_1: E(X) \neq E(Y);$$

we choose the level $1 - \alpha = 0.95 = 95\%$.

This means that we assume, based on experience, the equality of the variances and now take two samples of size $n = m = 15$ from the production of the first and second machines. These samples yield the following values: $\overline{X} = 19.800$ mm, $s_n^2 = 0.1234$, $\overline{Y} = 20.128$ mm, $s_m = 0.1362$ with $n = m = 15$. For the test statistic we obtain

$$t_{28} = \frac{0.328}{\sqrt{14 \cdot 0.1234 + 14 \cdot 0.1362}} \sqrt{\frac{15 \cdot 15 \cdot 28}{30}} = 2.49.$$

In Table VIII we find that at the 95% level, with $n + m - 2 = 28$, the acceptance region is the interval $(-2.048;\ 2.048)$. Therefore we reject the null hypothesis.

Finally, we turn to the *comparison of two samples in the case of normally distributed random variables if the unknown variances differ*. Below we give an approximative solution of this problem, due to Welch (see, e.g., [A 28], p. 146).

Making use of the above notation we compute from the two samples the value of the test statistic

$$t_f = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_n^2}{n} + \dfrac{s_m^2}{m}}}\,.$$

In the case of the validity of H_0 the distribution of t_f is close to the Student distribution with parameter f, the value of which is to be determined in the following way:

Let C be:

$$C = \frac{\dfrac{s_m^2}{m}}{\dfrac{s_n^2}{n} + \dfrac{s_m^2}{m}} < 1 .$$

Then

$$\frac{1}{f} = \frac{C^2}{m-1} + \frac{(1-C)^2}{n-1} .$$

With the parameter f follows the same procedure used in the two-sample Student test given before. (If f is not an integer then interpolation is necessary.)

4.3.6.5 Comparison of the Variances of More Than Two Gaussian Random Variables; the Bartlett Test. Let $X^{(1)}, X^{(2)}, \ldots X^{(r)}$ be normally distributed random variables. (E.g., some measurable property of products produced on r different machines). The equality of the variances, i.e., the null hypothesis

$$H_0\colon D^2(X^{(1)}) = D^2(X^{(2)}) = \ldots = D^2(X^{(r)})$$

is to be tested. The alternative will be H_1: there exist at least two indices j, k for which

$$D^2(X^{(j)}) \neq D^2(X^{(k)}) .$$

In order to make a decision between the two statements, take samples of sizes $n_1, n_2, \ldots, n_r$ of the variables $X^{(1)}, X^{(2)}, \ldots, X^{(r)}$. Let us denote by $s_{n_1}^2, s_{n_2}^2, \ldots, s_{n_r}^2$ the empirical variances of the corresponding samples and introduce the notation

$$\sum_{i=1}^{r} n_i = n, \quad s_n^2 = \frac{1}{n-r} \sum_{i=1}^{r} (n_i - 1)\, s_{n_i}^2 .$$

The following test statistic will be applied:

$$B^2 = \frac{2.3026}{C} \left((n-r) \log s_n^2 - \sum_{i=1}^{r} (n_i - 1) \log s_{n_i}^2 \right) ,$$

where

$$C = 1 + \frac{1}{3(r-1)} \left(\sum_{i=1}^{r} \frac{1}{n_i - 1} - \frac{1}{n-r} \right)$$

and log denotes the logarithm to the base 10. As Bartlett has proved, if $n_i \geqslant 4$ for $i = 1, 2, \ldots, r$, then the distribution of the statistic B^2 is close to the χ^2 distribution with parameter $n - r$. The critical region at level $1 - \alpha$ will be

$$\mathfrak{X}_c = \{B^2 \geqslant \chi^2_{n-r;1-\alpha}\},$$

where the critical value $\chi^2_{n-r,1-\alpha}$ is determined by the following relation:

$$P(\chi^2_{n-r} \leqslant \chi^2_{n-r,1-\alpha}) = 1 - \alpha.$$

Here the quantity χ^2_{n-1} stands for a random variable distributed according to the χ^2 distribution with parameter $n - r$.

The constant C is always larger than 1; therefore it is often convenient first to compare the value of CB^2 with the critical value. If we find that even CB^2 is smaller than $\chi^2_{n-r;1-\alpha}$, then the null hypothesis can be accepted, since in this case the value of $B^2 < CB^2$ is certainly less than $\chi^2_{n-r;1-\varepsilon}$. In the case of $CB^2 > X_{n-r;1-\alpha}$ the value of C must be determined and B^2 compared with the critical value.

For equal sample sizes, the expression for B^2 is somewhat simpler. Denoting by n_0 the common sample size, the test statistic has the form

$$B^2 = \frac{2.3026}{C} r(n_0 - 1) \left(\log s_n^2 - \frac{1}{r} \sum_{i=1}^{r} \log s_{n_i}^2\right).$$

The critical values $\chi^f_{2,n-r}$ of the chi-square distribution are given in Table 9 (p.398). An example will be given later.

4.3.6.6 Comparison of the Expected Values of More Than Two Gaussian Random Variables. Let $X^{(1)}, X^{(2)}, \ldots, X^{(r)}$ be normally distributed random variables with common variances: $D^2(X^{(i)}) = \sigma^2$, $i = 1, 2, \ldots, r$. For testing the null hypothesis

$$H_0\colon E(X^{(i)}) = a_i = a, \quad i = 1, 2, \ldots r$$

against the alternative

$$H_1\colon \text{there exist at least two indices } j \text{ and } k \text{ with } a_j \neq a_k,$$

the following procedure will be used.

Let $X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}$ be a sample of size n_i taken on the ith random variable $X^{(i)}$, $i = 1, 2, \ldots, r$. Further, let $X^{(i)}$ be the mean of this sample and let $n = \sum_{i=1}^{r} n_i$.

We construct the following two sums of squares:

$$(r - 1)\, s_1^2 = \sum_{i=1}^{r} n_i (\overline{X}^{(i)} - \overline{X})^2$$

with

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{r} n_i \overline{X}^{(i)} \tag{28}$$

and

$$(n - r) s_2^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_j^{(i)} - \overline{X}^{(i)})^2 . \tag{29}$$

It can be shown that under the validity of the null hypothesis both $\frac{(r-1)s_1^2}{\sigma^2}$ and $\frac{(n-r)s_2^2}{\sigma^2}$ are distributed according to the χ^2 distribution with parameters $r - 1$ and $n - r$, respectively; furthermore, they are independent. On the other hand, if H_1 is true, then $\frac{(n-r)s_2^2}{\sigma^2}$ still has a χ^2 distribution with parameter $n - r$, whereas the distribution of $\frac{(r-1)s_1^2}{\sigma^2}$ differs from the χ^2 distribution. As it can be seen from expression (29), the expected value of each $X_j^{(i)} - \overline{X}^{(i)}$ vanishes: these are differences where both members have the same expectation. But the expected value of $\overline{X}^{(i)} - \overline{X}$ is $a_i - \overline{a}$ with $\overline{a} = \frac{1}{n} \sum_{i=1}^{r} n_i a_i$. The difference $a_i - \overline{a}$ can be zero for each i only if the null hypothesis holds; these are quantities which are influenced by the differences in the expectations of the (different) samples.

According to Section 2.9.6, the relation

$$E(s_1^2) = E(s_2^2) = \sigma^2$$

holds for the expected values of the given χ^2-distributed variables under H_0, whereas, as can easily be seen, the hypothesis H_1 has the consequence (which can be considered as alternative)

$$H_1': \; E(s_1^2) > E(s_2^2).$$

For making a decision between H_0 and H_1, the following test statistic will now be applied:

$$F = \frac{s_1^2}{s_2^2} .$$

The procedure is the same as that given in 4.3.6.3 for the F test in the one-sided case. If $s_1^2 < s_2^2$ holds, then H_0 will be accepted. Otherwise we have at level $1 - \alpha$ the critical region

$$\mathfrak{X}_c = \{F > F_{f_1, f_2; 1-\alpha}\} \text{ with } f_1 = r - 1, \; f_2 = n - r.$$

We turn now to an example from the textile industry. (This example is taken from [A 46]).

Five different lots of artificial silk are to be compared with respect to the number of twists occurring in a thread of $\frac{1}{2}$ meter length. Denoting by $X^{(i)}$ $(i = 1, 2, 3, 4, 5)$ the quantity observed in the case of the ith lot, the homogeneity of the lots will mean the equality of their variances and expected values.

First, we consider the null hypothesis H_0: $D^2(X^{(i)}) =$ constant $(i = 1, 2, 3, 4, 5)$. On the basis of earlier experiences we assume the normality of the random variables; therefore in testing H_0 the Bartlett test can be applied. If H_0 is accepted, we go on with the examination of the hypothesis

$$H_0'\colon\ E(X^{(i)}) = \text{constant}\ (i = 1, 2, 3, 4, 5).$$

The samples of size 12 taken from the lots resulted in the figures given in Table 6b.

TABLE 6b

j \ i	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$X^{(4)}$	$X^{(5)}$
1	19.4	18.4	18.6	20.3	22.4
2	18.6	22.4	20.6	16.8	22.0
3	21.2	23.0	20.1	16.7	19.3
4	16.8	19.4	20.1	23.0	22.7
5	18.6	22.6	22.9	20.3	17.4
6	18.0	22.2	18.5	23.0	19.3
7	21.6	21.2	18.8	20.7	19.9
8	17.8	20.8	21.3	20.9	20.0
9	16.8	21.6	16.9	23.2	10.5
10	20.4	20.2	19.2	21.5	19.4
11	19.2	22.2	21.9	19.3	19.3
12	16.4	20.4	17.9	18.4	17.5
$X^{(i)}$	18.73	21.20	19.73	20.34	19.89
s_i^2	3.06	1.97	3.20	5.05	2.98

The last two rows contain the sample means and empirical variances: $\bar{X}^{(i)}$ and s_i^2 $(i = 1, 2, 3, 4, 5)$.

We will decide on H_0 at the level 95%; $r - 1$ being 4, the test statistic CB^2 is to be compared with the critical value $\chi^2_{4;\,0.95} = 9.49$.

Since the common sample size equals 12, the test statistic has the form:

$$CB^2 = 2.3026(55 \log s^2 - 11 \sum_{i=1}^{5} \log s_i^2),$$

where

$$s^2 = \frac{1}{4}\sum_{i=1}^{5} s_i^2, \quad C = 1 + \frac{2}{55}.$$

Substituting the numerical values, we find (Table IX, p. 398) that $CB^2 = 2.4847 < \chi^2_{4;\,0.95} = 9.49$. Therefore, we accept the equality of the variances.

For making a decision concerning the hypothesis H_0', again at the level $1 - \alpha = 0.95$, we compute the sums of squares:

$$s_1^2 = \frac{1}{4}\sum_{i=1}^{5} 12\,(\bar{X}^{(i)} - \bar{X})^2 = 9.7533$$

and

$$s_2^2 = \frac{1}{55}\sum_{i=1}^{5}\sum_{j=1}^{12}(X_j^{(i)} - \bar{X}^{(i)})^2 = 3.252,$$

their quotient being

$$F = \frac{s_1^2}{s_2^2} = 2.999.$$

In Table VII, corresponding to the pair of parameters (4,55), we find $F_{4,55;0.95} = 2.54 < F = 2.999$. Consequently, at the level 95%, we reject the null hypothesis concerning the homogeneity of the lots: the expected values show, at the given level, significant differences.

4.3.7 Tests of 'Goodness of Fit'. Tests for Homogeneity and Independence

4.3.7.1 Inference Concerning the Distribution Function of a Random Variable. In the preceding section test procedures were given for making decisions concerning *normally distributed* random variables. Normality was always a part of both H_0 and H_1. The question of how to verify this assumption will now be raised. More generally, we shall treat the problem of making inferences about the distribution function of a random variable, and of comparing samples in order to decide whether or not they come from the same distribution.

Suppose that in practice we meet a certain random variable for the first time. By considering the physical, chemical or technical background, we are sometimes able to obtain information on the type of distribution from theoretical considerations. We refer here to Section 2.9, where we mentioned different circumstances which bring about certain kinds of distributions (normal, exponential, Poisson etc.). In such a case our next step is to compare our assumption with practical experience. A sample of the random variable in question is taken in order to examine how "well" the sample fits the supposed distribution or a certain element of the supposed type of distribution.

In other cases we have no hint at all as to the type of distribution except general properties such as whether it is discrete or continuous, bounded or unbounded etc. Here, large samples (e.g. of size 100—350) are needed. The shape of the frequency diagram or density diagram of the sample may indicate how the values are distributed on the real line. From the figure obtained, conclusions can be drawn in many cases as to the type of the unknown distribution. Our next task is again to use statistical methods for the verification of the assumption thus established.

Procedures which help to determine whether a sample of a random variable is of a given distribution or a given type of distribution are called *tests for goodness of fit*. The Kolmogorov test and the so-called chi-square (χ^2) test have further important applications (e.g. testing the independence of two random variables), which will be treated too. The test of Kolmogorov can be used for continuous variables only, while the χ^2 test applies to both the discrete and the continuous case. For both tests the distribution of the test statistic under the null hypothesis is independent of the underlying distribution in the null hypothesis: these are the so-called *distribution-free* tests.

We shall consider, in addition to these tests of general applicability, special tests devised for the normal and exponential case. For testing normality, we shall describe a simple graphical method.

Finally, a method based on order statistics will be given for comparing two samples to determine whether or not they come from the same distribution.

4.3.7.2 The Test of Kolmogorov. Let $X_1, X_2, \ldots, X_n$ be a sample of size n, the elements of which are observations on a continuous random variable X. Further, let $F(x)$ be a continuous distribution function, fully specified. For testing the null hypothesis

$$H_0\colon\ P(X \leqslant x) = F(x)$$

against the two-sided alternative

$$H_1\colon\ P(X \leqslant x) = G(x) \not\equiv F(x)$$

the following test statistic is used:

$$D_n = \sqrt{n} \sup_{(x)} | F_n(x) - F(x) | ,$$

where $F_n(x)$ is the empirical distribution function belonging to the given sample. For the determination of D_n the values $| F_n(x_i - 0) - F(x_i) |$ and $| F_n(x_i) - F(x_i) |$ are to be computed ($i = 1, 2, \ldots, n$).

In Section 3.3.3 the asymptotic distribution function of the test statistic under H_0 can be found. The critical region, at the level $1 - \alpha$, has the form

$$\mathfrak{X}_c = \left\{ D_n \geqslant D_{1-\alpha} \right\}.$$

The values of $D_{0.95}$ and $D_{0.99}$ are tabulated in Table X (p. 399).

4.3.7.3 Testing for Normality. We first give a graphical method for examining whether or not a large sample comes from a normal distribution. Though this is not an exact method, it is a useful and simple aid in getting preliminary information about the normality of the distribution in question. In extreme cases, however, we may come to a final decision if the data show either a very good fit to normality or if the results definitely contradict the normal assumption. In intermediate cases a more rigorous method is necessary.

For large samples this may be the χ^2 test. For small samples we will give another test.

The null hypothesis is:

H_0: the random variable X is distributed according to the normal law: $P(X \leqslant x) = \Phi(x;\ \mu, \sigma)$ with, perhaps unknown, parameters μ and σ.

The alternative hypothesis is:

H_1: the distribution of the random variable X differs from the Gaussian one.

Our aim is to make a decision between H_0 and H_1 based on the sample $X_1, X_2, \ldots, X_n$.

The graphical method. This is a heuristic method in most cases, suitable only for a first rough examination; in connection with it we cannot speak about a level at all. We plot the grouped data on a so-called normal probability paper (grid) and then we judge by sight whether the figure obtained lies close to some straight line or whether it deviates more than a random amount from any straight line. In the first case we accept H_0; in the second case we reject it.

The probability paper has a horizontal and a vertical scale. The former is divided up uniformly; here the possible values of the random variable are to be marked. The vertical scale shows the respective probabilities (see Fig. 30). The principle of scaling is the following: the value 0.5 is signed to the center. Choosing a distance d, at a distance $+d$ and $-d$ from 0.5 on the vertical scale, $\Phi(+1) = 0.8413$ and $\Phi(-1) = 0.1587$, at the distance $\pm 2d$ from 0.5 the values $\Phi(2) = 0.9772$ and $\Phi(-2) = = 0.0228$ will be marked etc.

Let us now group the data into r classes with endpoints $-\infty = y_0 < y_1 < y_2 < \dots < y_{r-1} < y_r = \infty$ and plot on the grid the points $\left(y_i, \sum_{j=1}^{i} g_j\right)$, $i = 1, 2, \dots, r-1$, where g_j denotes the relative frequency of the sample elements lying within the interval (y_{j-1}, y_j). If the sample comes from a Gaussian distribution, the points obtained lie close to a straight line.

Figure 30 shows three lines. The one marked (a) can be accepted at sight as corresponding to the normal law; the points belong to the line; (b) shows a significant deviation from the Gaussian distribution; (c) does not allow, at sight, a reassuring judgement concerning normality: some other, more objective, procedure is needed.

The method described next is advisable in the small sample case; it is a transformation of the sample so that a small sample test can be applied.

A test for normality based on a transformation of the sample elements. [B 67] This procedure is as follows: We transform the original sample elements $X_1, X_2, \dots, X_n$ into a new "sample" of size $n - 2$, i.e. we reduce the original sample size by 2. Denoting the new sample by $Y_1, Y_2, \dots, Y_{n-2}$, it is known that under H_0, the transformed sample consists of independent elements distributed according to $N(0, 1)$. Consequently, the null hypothesis

$$H_0\colon\ P(Y_i \leqslant x) = \Phi(x;\ 0, 1) = \Phi(x), \quad i = 1, 2, \dots, n-2$$

can be tested, e.g., with the aid of the Kolmogorov test (Section 4.3.7.5).

The transformation formulas are the following:

$$Y_i = \frac{X_i - \overline{X}'}{s}\, \psi_{n-2}\left(\frac{|X_{n-1} - X_n|}{\sqrt{s}}\right), \quad i = 1, 2, \dots, n-2$$

Denoting by $\overline{X}$ the sample mean belonging to the original sample, $\overline{X}'$ has the form

$$\overline{X}' = \frac{\overline{X} + \sqrt{\frac{n}{2}}\,(X_{n-1} + X_n)}{n + \sqrt{2n}},$$

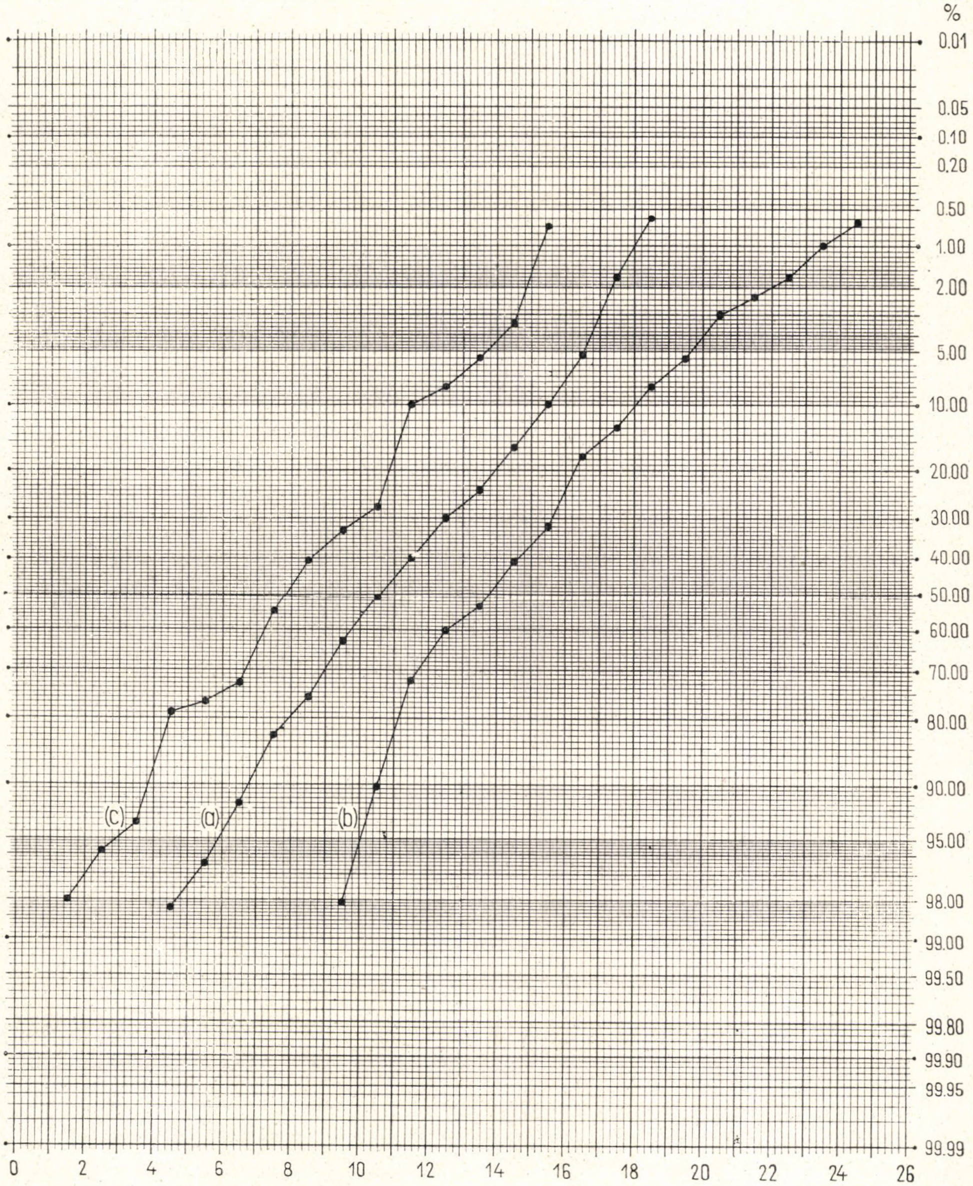


Fig. 30. Graphical representation of statistical data (large samples) in a Gaussian lattice: (a) normality seems to be acceptable; (b) normality does not seem to be present; (c) further investigation is needed

while

$$s^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n} X_i^2 - n\,\bar{X}^2 - \frac{1}{2}(X_{n-1} - X_n)^2\right].$$

The values of the function $\psi_{n-2}(t)$ can be obtained with the aid of the Student and χ^2 distributions functions. Let us denote by $S(t|r)$ the distribution function of the Student distribution with parameter r (see Section 2.9.8) and by $H(t|r)$ the distribution function of the χ^2 variable with parameter r (see Section 2.9.6). The definition of the function $\psi_{n-2}(t)$ is given by means of the following relation:

$$1 - H([\psi_{n-2}(t)]^2 \mid n-2) = 2\,S(t \mid n-2) - 1, \qquad t \geqslant 0\ \psi_{n-2}(t) > 0.$$

As the tabulation of the Student and χ^2 distributions given at the end of our book is insufficient for the computation of the values of $\psi_{n-2}(t)$, we refer to the Biometrika Tables for Statisticians [C 8] or to the Tables given by Bolshev and Smirnov [C1].

4.3.7.4 Testing the Hypothesis of Exponentiality (Störmer's Test). Let $X_1, X_2, \ldots, X_n$ be n independent, identically-distributed random variables. The hypothesis to be tested is whether these variables are exponentially-distributed with unknown parameter λ:

$$H_0\colon\ P(X_i < x) = 1 - e^{-\lambda x},\ x \geqslant 0, \quad i = 1, 2, \ldots, n.$$

Let us denote by $X_0^*, X_1^*, \ldots, X_n^*$ the ordered elements with $X_0^* = 0$. We introduce now the new ordered variables

$$Y_i^* = \frac{\sum\limits_{j=0}^{i-1} (n-j)(X_{j+1}^* - X_j^*)}{\sum\limits_{j=1}^{n} X_j^*}, \quad i = 1, 2, \ldots, n-1.$$

If the assumption of exponentiality is valid, then the random variables: $Y_1^* < Y_2^* < \ldots < Y_{n-1}^*$ can be regarded as the ordered form of a sample of size $n-1$ taken on a random variable Y distributed according to the uniform distribution (see Section 2.9.11). Denoting by $Y_1, Y_2, \ldots, Y_{n-1}$ the (unknown) independent sample elements with the given ordered form $Y_1^*, Y_2^*, \ldots, Y_{n-1}^*$, the following null hypothesis is to be tested:

$$P(Y_i \leqslant x) = U(x) = \begin{cases} 0, & \text{if } x \leqslant 0 \\ x, & \text{if } 0 < x \leqslant 1. \\ 1, & \text{if } 1 < x. \end{cases}$$

This hypothesis can be examined with the aid of the Kolmogorov test (Section 4.3.7.2), the latter being based on the *ordered* sample elements (i.e. the Y_i^*'s) only.

4.3.7.5 Comparison of Two Samples (The Kolmogorov–Smirnov Two-Sample Test). As repeatedly mentioned, in practice it is very often necessary to examine whether or not two samples come from the same distribution. The comparison of a certain measure of two quality characteristics in two lots of two production procedures may lead to this problem. Sometimes the one-sided version is of practical importance: e.g., has the life span (or another measure) of certain parts really increased by introducing a new procedure? Is it larger in lot A than in lot B? We must emphasize that the method treated here can be applied only for *random variables with continuous distribution functions.*

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be samples taken from the continuous random variables X and Y. Using $F(x)$ and $G(x)$ for the continuous distribution functions of X and Y, the null hypothesis

$$H_0\colon\ F(x) \equiv G(x)$$

is to be tested; the two-sided alternative is

$$H_1\colon\ F(x) \not\equiv G(x),$$

while the one-sided alternative is

$$H_1^+\colon\ F(x) \geqslant G(x) \text{ and } F(x_0) > G(x_0) \text{ for some } x_0.$$

Against H_1 the test statistic is

$$D_{n,m} = \max_{(x)} \mid F_n(x) - G_m(x) \mid,$$

while against H_1^+ the one-sided maximal deviation

$$D_{n,m}^+ = \max_{(x)} (F_n(x) - G_m(x))$$

will be used. $F_n(x)$ and $G_m(x)$ denote the empirical distribution functions belonging to the first and second samples. As critical region for testing against H_1 at level $1 - \alpha$ we have to choose

$$\mathfrak{X}_c = \{D_{n,m} > D_\alpha\}$$

and for testing against H_1^+,

$$\mathfrak{X}_c^+ = \{D_{n,m}^+ > D_\alpha^+\}$$

where the critical values D_α and D_α^+ are determined by the following relations

$$P(D_{n,m} \leqslant D_\alpha \mid H_0) = 1 - \alpha, \qquad P(D_{n,m}^+ \leqslant D_\alpha^+ \mid H_0) = 1 - \alpha.$$

Table XI (p. 399) gives the critical values for the case of equal sample sizes $m = n$ at the levels of 95% and 99%. The Table contains the integer values nD_α and nD_α^+, the values $n\,D_{n,n}$ and $n\,D_{n,n}^+$ are to be compared with them.

Having equal sample sizes for the determination of the test statistics, the following simple procedure can be applied. Let us arrange the sample elements according to their order of magnitude and write down the second sample under the first:

$$X_1^*,\ X_2^*, \ldots, X_n^*,$$

$$Y_1^*,\ Y_2^*, \ldots, Y_n^*.$$

Now, if each X_i^* is larger than the corresponding Y_i^* then $nD_{n,n}^+ = 0$; if this is not the case, then let us shift the first row to the left with one or more steps until each X_i^* becomes larger than the Y_j^* below it. If k steps are needed, k is the smallest integer for which $X_{i+k}^* > Y_i^*$ for $i = 1, 2, \ldots, n - k$, then $nD_{n,n}^+ = k$. Changing the role of the two samples, this means shifting the first sample to the right till each X^* is smaller than the respective $Y_{i+k'}^*$, $(i = 1, 2, \ldots, n - k')$. Then

$$k' = n \min_{(x)} (F_n(x) - G_n(x)) = nD_{n,n}^- \text{ and } nD_{n,n} = \max (k, k').$$

If, e.g., $X_1 = 5$, $X_2 = 4$, $X_3 = 3$, $Y_1 = 2$, $Y_2 = 6$, $Y_3 = 1$, then we have

$$\begin{array}{ccc} 3 & 4 & 5 \\ 1 & 2 & 6; \end{array}$$

shifting the first row by one step to the right,

$$\begin{array}{cccc} 3 & 4 & 5 & \\ & 1 & 2 & 6 \end{array}$$

leads to the value $k = 1$; on the other hand, shifting to the opposite direction:

$$\begin{array}{ccccc} & & 3 & 4 & 5 \\ 1 & 2 & 6 & & \end{array}$$

$k' = 2$ steps are needed. This means that $3D_{3,3}^+ = 1$ and $3D_{3,3}^- = 2$, so $3D_{3,3} = \max(1, 2) = 2$.

4.3.7.6 The χ^2 Test; Its Application for Testing of Goodness of Fit, of Homogeneity and of Independence. This test is based on the following theorem: Let $A_1, A_2, \ldots, A_r$ be the possible outcomes of an experiment, a complete system of events with probabilities $p_1, p_2, \ldots, p_r, \sum_{i=1}^{r} p_i = 1$. Let us make observations on the phenomenon in a large number of cases, say N times. Denoting by n_i the frequency of A_i, obviously $\sum_{i=1}^{r} n_i = N$ must be true. Constructing now the statistic

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - Np_i)^2}{Np_i},$$

its distribution is close to the χ^2 distribution with parameter $r - 1$, if N is large enough. The null hypothesis is the assumption that the true values of the probability of A_i agree with p_i for each i; if this is not true, the distribution of the statistic will differ from the χ^2 distribution and its value will exceed the critical value with high probability.

For testing whether a sample fits a given distribution, we will proceed as follows: If the random variable is discrete, then A_i means that the random variable X takes on a certain value x_i: $A_i = \{X = x_i\}$ and n_i is the number of sample elements taking on this value in the totality of N observations. If the random variable has a continuous distribution, then the sample taken on X will be grouped by dividing the real axis into intervals. Having such an interval (y_{i-1}, y_i) for the event A_i, we have $A_i = \{y_{i-1} < X \leqslant y_i\}$ and n_i will be the number of sample elements falling into the interval (y_{i-1}, y_i). The corresponding probabilities will depend on the distribution $F(x)$ assumed in the null hypothesis:

$$P(A_i) = F(y_i) - F(y_{i-1}).$$

Having a large sample, the events A_i must be chosen in such a way that the corresponding frequencies have values of at least 6—10. For this reason, in the case of a discrete distribution, if low frequencies occur, grouping is advisable.

In the case of continuous random variables, the formation of longer intervals at both ends may be necessary.

In the two-sample case the exact probabilities p_i do not play a role: only the relative frequencies will be used.

We turn now to the different possibilities of application.

Testing goodness of fit. Let X be a random variable, discrete or continuous, and let us take a sample of size N on from it: $X_1, X_2, \ldots, X_N$. The null hypothesis is

$$H_0\colon\ P(X \leqslant x) = F(x),$$

where the distribution function $F(x)$ is given, its type being known. Certain parameters may or may not be known.

For the alternative we have

$$H_1\colon\ P(X \leqslant x) = G(x) \not\equiv F(x).$$

This means that we have to decide whether or not the distribution function of X agrees with $F(x)$.

If the range of the distribution is (a, b) where a and b can take finite or infinite values, we divide it into r parts with the aid of points $a = y_0 < y_1 < y_2 < \cdots < y_{r-1} < y_r = b$. Let us denote by n_i the numbers of X_j's in the interval (y_{i-1}, y_i) $(i = 1, 2, \ldots, r)$. Concerning the choice of r we can say that if $N \sim$ 80–30, r can be taken as 8–15. The values y_i can be chosen by taking into account the distribution function $F(x)$; we do not want the differences in the probabilities of the different intervals to be too big. These are not general rules; for further considerations of the use of the χ^2 test we refer to the book by Lehmann [A 30] and to the references given there.

Next follows the computation of the probabilities:

$$p_i = F(y_i) - F(y_{i-1}).$$

E.g., if $F(x) = \Phi(x;\, \mu_0, \sigma_0)$ with known expected value μ_0 and known variance σ_0, then

$$p_i = \Phi(y_i, \mu_0, \sigma_0) - \Phi(y_{i-1};\, \mu_0, \sigma_0) = \Phi\left(\frac{y_i - \mu_0}{\sigma_0}\right) - \Phi\left(\frac{y_{i-1} - \mu_0}{\sigma_0}\right).$$

If the parameters are unknown they will be estimated from the sample: $\mu_0 \sim \overline{X}$ and $\sigma_0^2 \sim s_N^2$, i.e., the sample mean and empirical variance will be used:

$$\hat{p}_i = \Phi\left(\frac{y_i - \overline{X}}{s_N}\right) - \Phi\left(\frac{y_{i-1} - \overline{X}}{s_N}\right), \quad i = 1, 2, \ldots, r.$$

The notation $\hat{p}_i$ indicates the fact that the probabilities are estimated; this *influences the parameter of the test statistic*. Under the original assumptions, the χ^2 statistic has χ^2 parameter $r - 1$, but if in $F(x)$ (and consequently in the p_i) the number of estimated parameters is s, then the test statistic is approximately χ^2 distributed with parameter $r-1-s$. In the

Gaussian case, i.e., in the example just given, the parameter value will be

$$r - 1 - 2 = r - 3.$$

The next step is the computation of the test statistic

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - Np_i)^2}{Np_i}$$

with parameter $r - 1$, or in the estimated case

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - N\hat{p}_i)^2}{N\hat{p}_i}$$

with parameter $r - 1 - s$, where s is the number of parameters in $F(x)$ estimated from the sample.

Finally, the critical region at level $1 - \alpha$:

$$\mathfrak{X}_c = \{\chi^2 > \chi^2_{f, 1-\alpha}\}$$

where $f = r - 1$, or $r - 1 - s$ depending on whether the distribution is fully specified by the null hypothesis, or whether s parameters have been estimated from the sample. The critical values $\chi^2_{f;1-\alpha}$ are tabulated in Table IX, p. 398.

Test for homogeneity. Two random variables X and Y are to be compared by means of samples $X_1, X_2, \ldots, X_N$ and $Y_1, Y_2, \ldots, Y_M$ taken from them. We are to decide whether their distribution functions agree. The null hypothesis is

$$H_0\colon\ P(X \leqslant x) = P(Y \leqslant x), \quad -\infty < x < \infty.$$

We did not use any notation for the distribution functions because they do not play any role at all in the test procedure (unless the power function is to be determined; for this topic see Birnbaum [B 8], Chapman [B 16], Vincze [B 82]).

First we divide the interval (a, b), which contains both samples, into r intervals at the points $a = y_0 < y_1 < y_2 < \ldots < y_{r-1} < y_r = b$. Let us denote by n_i and m_i the number of elements in the first and second sample contained in the interval (y_{i-1}, y_i), $i = 1, 2, \ldots, r$. Then as test statistic the following expression will be used:

$$\chi^2 = MN \sum_{i=1}^{r} \frac{\left(\frac{n_i}{N} - \frac{m_i}{M}\right)^2}{n_i + m_i},$$

which has a distribution close to the χ^2 distribution with parameter $r - 1$ under the null hypothesis. If we find an observed value of χ^2 larger than the critical value $\chi^2_{r-1;1-\alpha}$, then we reject the null hypothesis at level $1 - \alpha$ and do not consider the two samples as coming from the same population.

Testing the independence of two random variables. Let X and Y be two random variables occurring in the same experiment; that is they form a vector variable (X, Y). We wish to test whether X and Y can be considered as independent variables:

$$H_0\colon\ P(X \leqslant x,\ Y \leqslant y) = P(X \leqslant x)P(Y \leqslant y)$$

for each pair of real numbers (x, y). The two variables may be different in character: the one may be continuous and the other discrete, they may be of different signs, etc. For example, we may examine the independence of the diameter and breaking strength of a cylinder, the number of yarn breakings and the quantity produced on a textile machine, etc.

The alternative is, in this case,

$$H_1\colon\ P(X \leqslant x,\ Y \leqslant y) \neq P(X \leqslant x)\ P(Y \leqslant y)$$

at least for certain values of x and y.

To decide between H_0 and H_1 we take a sample from (X, Y): (X_i, Y_i) $i = 1, 2, \ldots, N$. We emphasize that X_i and Y_i are related variables; they belong to and will be observed in the same experiment.

Let us suppose that the X_i's are contained in the interval (a, b) while the Y_i's are contained in the interval (c, d). Let us divide (a, b) into r parts by means of the points $a = x_0 < x_1 < x_2 < \ldots < x_{r-1} < x_r = b$, and similarly (c, d) into s parts at the values $c = y_0 < y_1 < y_2 < \ldots < y_{r-1} < y_r = d$. Thus we obtain $r - s$ cells for the two-dimensional sample elements. Let us denote by n_{kl} the number of sample elements (X_i, Y_i) with first coordinate in the kth interval (x_{l-1}, x_k) and at the same time with second coordinate Y_i in the lth interval (y_{i-1}, y_l), i.e., $n_{k,l}$ is the frequency of the event $\{x_{k-1} < X \leqslant x_k, y_{l-1} < Y \leqslant y_l\}$, $k = 1, 2, \ldots, r$, $l = 1, 2, \ldots, s$. Obviously, the sum $\sum\limits_{l=1}^{s} n_{k,l} = n_{k\cdot}$ gives the total number of sample elements having first coordinate in the interval (x_{k-1}, x_k), disregarding the second coordinate. Similarly, $\sum\limits_{k=1}^{r} n_{kl} = n_{\cdot l}$ is the frequency of the event $\{y_{l-1} < Y \leqslant y_l\}$, independent of the corresponding X. These are the frequencies belonging to the marginal distributions $F(x, \infty)$ and $F(\infty, y)$ respectively. Furthermore, $\sum\limits_{l=1}^{s} n_{\cdot l} = \sum\limits_{k=1}^{r} n_{k\cdot} = N$.

The test statistic is:

$$\chi^2 = N \sum_{k=1}^{r} \sum_{l=1}^{s} \frac{\left(n_{kl} - \dfrac{n_{k\cdot}n_{\cdot l}}{N}\right)^2}{n_{k\cdot}\, n_l},$$

which is, for large N, approximately χ^2-distributed with parameter $(r-1)(s-1)$. Thus the test procedure can be applied at a chosen level $1-\alpha$.

A very important and often used special case is $r = s = 2$, which corresponds to the question of whether or not two (alternative) properties are independent. Such a problem was treated in Section 1.8, and its solution will be given now. We choose the level $1-\alpha = 0.95$. In Table 3 the following figures appear: $n_{11} = 464$; $n_{12} = 4$; $n_{1\cdot} = 468$; $n_{21} = 6$; $n_{22} = 26$; $n_{2\cdot} = 32$; $n_{\cdot 1} = 470$; $n_{\cdot 2} = 30$; $N = 500$.

In this case the test statistic has the following simple form:

$$\chi^2 = N \frac{(n_{11}\, n_{22} - n_{12}\, n_{21})^2}{n_{1\cdot}\, n_{2\cdot}\, n_{\cdot 1}\, n_{\cdot 2}}$$

with parameter $(r-1)(s-1) = 1$.

Substituting the observed values given in the above table we obtain $\chi^2 = 343.20$. Because the critical value in Table IX is $\chi^2_{1;0.95} = 3.84$ and χ^2 is considerably larger than that, we reject the independence of the two properties at level 95% (even at level 99.95%).

In case the hypothesis of independence is rejected, a qualitative examination of the dependence existing between the two random variables is advisable; such an investigation may be performed by using the regression analysis treated in Section 4.4.2.

4.3.8 Application of Statistical Tests to the Control of Quality

Activities of this kind may be grouped into two main classes: 1. A systematic and permanent work (standard methods), the fundamentals of which will be presented in the third chapter of the present book. 2. The occasional application of various test procedures in special problems of quality control. Even if applying a standard statistical quality control method, a circumstance may arise which makes it necessary to use a special statistical examination not belonging to the original quality control method.

In any case the control of quality can be characterized, broadly speaking, as a comparison of the actual data with some desired quantities or qualities, i.e., with prescribed data, with standards or with actual data coming from other sources.

Consequently, test procedures play an important role in statistical quality control. The control of a production process is, in the majority of cases, nothing other than a routine for repeated application of a definite statistical test, repeated at well-defined or randomly chosen instants. Acceptance plans are also test procedures appropriate for routine application in the same or in similar situations.

In applying a test systematically, the OC function or other characteristics make it possible to be aware of the economic and other consequences of the procedure. We know that over a long period of time the statistical fluctuations will be balanced out, and statistical regularity will dominate. But, as mentioned before, statistical tests are sometimes used in special circumstances, without many repetitions. E.g., in a factory there may occur once, or very rarely, the acceptance of lots of exceedingly high value. Similarly, a scientific experiment may be carried out only once. In such circumstances statistical tests are used in a form modified slightly from that treated above.

If for example, testing for homogeneity is needed and the χ^2 test is used, the level $1-\alpha$ of the test is usually given in advance. This, together with the parameter of the χ^2 statistic, determines the critical value. However, applying the χ^2 test to a single case, the following version of the procedure is often used: first we determine the value of the test statistic and look in the table of the distribution function for the level at which it must be rejected. Suppose, e.g., that the actual value of the test statistic is $\chi^2 = 29.2$ and the parameter is 20; in Table IX we find that $\chi^2_{20,\,0.90} =$ $= 28.4$ and $\chi^2_{20;\,0.95} = 31.4$. This means that we accept the null hypothesis at level 95% but we reject it at level 90%. Being mainly interested in the quality (the homogeneity) we prefer, in such cases, to accept the null hypothesis at a low level, e.g., at a level smaller than 80% rather than at a higher level. In any case, we prefer finding an actual value of the test statistic closer to its expected value, which is 20 in the present example. If we cannot accept the result, a second sample should be chosen.

Turning to the routine application of a test, the above consideration can be utilized in the following way: the test statistic, being a random variable, fluctuates under a simple null hypothesis according to a given distribution law; the fluctuation occurs around its expected value. If we find that the observed value of the test statistic very often lies near the critical value, though still inside the acceptance region, we may conclude that the true distribution does not agree with that given by H_0. The control chart method used in the control of production processes is a very good tool for keeping an eye on the behavior of the test statistic.

The points made here indicate that recording the sampling results is strongly advisable in the case of acceptance plans as well. We will return to this problem later.

4.3.9 Example for the Computation of Error Probabilities

Let us suppose that a certain measurement of a part is to be controlled during the production process. Using X to denote the measurement as a random variable, it is known by earlier experience that X is distributed according to the normal law with the prescribed and known expected value μ_0 and variance σ_0^2. This coincides with our null hypothesis:

$$H_0\colon\ P(X \leqslant x) = \Phi(x;\ \mu_0, \sigma_0)$$

or

$$E(X) = \mu_0,\ D^2(X) = \sigma_0^2.$$

In order to check the production process we take a sample containing four pieces of the lot produced by the machine every half an hour. The four pieces are chosen randomly from the production of the last period. Let X_1, X_2, X_3, X_4 be the sample elements; from these we compute the sample mean and variance

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4}{4}, \qquad s_4^2 = \frac{1}{3}\sum_{i=1}^{4}(X_i - \overline{X})^2 .$$

As $\overline{X}$ is distributed according to $N\left(\mu_0, \frac{\sigma_0^2}{4}\right)$ it will lie, if H_0 is true, in the interval

$$\left(\mu_0 - 2\frac{\sigma_0}{2} = \mu_0 - \sigma_0,\quad \mu_0 + 2\frac{\sigma_0}{2} = \mu_0 + \sigma_0\right)$$

with probability close to 95%.

Similarly, we know that $(n-1)\dfrac{s_n^2}{\sigma_0^2} = 3\dfrac{s_4^2}{\sigma_0^2}$ is distributed according to the χ^2 distribution with parameter 3. Consequently, using Table IX (p. 398) the following relation holds under the validity of H_0:

$$P\left(3\frac{s_4^2}{\sigma_0^2} \leqslant 7.81 \mid H_0\right) = 0.95 .$$

Let us suppose now that we choose as the acceptance region in our control process the following domain:

$$\left\{\mu_0 - \sigma_0 < \overline{X} \leqslant \mu_0 + \sigma_0, \quad s_4^2 \leqslant \frac{1}{3} 7.81\, \sigma_0^2\right\}.$$

This means that we reject the null hypothesis if either $\overline{X}$ lies outside of the interval $(\mu_0 - \sigma_0, \mu_0 + \sigma_0)$ or if the observed value of s_4^2 turns out larger than $2.60\, \sigma_0^2$ (or if both events occur). Using the fact that in the case of a Gaussian distribution the sample mean and sample variance are independent random variables, the probability of the error of first kind can be computed in the following way:

$$\begin{aligned} 1 &- P(|\overline{X} - \mu_0| \leqslant \sigma_0, s_4^2 \leqslant 2.60\, \sigma_0^2 \mid H_0) \\ &= 1 - P(|\overline{X} - \mu_0| \leqslant \sigma_0 \mid H_0)\; P(s_4^2 \leqslant 2.60\, \sigma_0^2 \mid H_0) \\ &= 1 - 0.954 \cdot 0.95 = 0.0937 = 9.37\%. \end{aligned}$$

As an example of the determination of the probability of an error of the second kind let us suppose an increase of $\frac{3}{2}\sigma_0$ in μ_0, and a simultaneous increase in the variance to $\frac{16}{9}\sigma_0^2$, make an intervention in the production process necessary. We define this event to be H_1:

$$H_1: \; E(X) = \mu_0 + \frac{3}{2}\sigma_0 \quad D^2(X) = \frac{16}{9}\sigma_0^2.$$

Thus, the probability of an error of the second kind will be obtained from the probability of the acceptance region, when that alternative is true.

As the two statistics are independent under the alternative hypothesis as well, the corresponding probabilities for the sample mean and variance can be calculated separately.

For the sample mean we have

$$\begin{aligned} &P(\mu_0 - \sigma_0 < \overline{X} \leqslant \mu_0 + \sigma_0 \mid H_1) = \\ &= P\left(-\frac{15}{4} < \frac{\overline{X} - \left(\mu_0 + \frac{3}{2}\sigma_0\right)}{\frac{4}{3}\sigma_0} \leqslant -\frac{3}{4} \,\middle|\, H_1\right) = \\ &= \Phi\left(-\frac{3}{4}\right) - \Phi\left(-\frac{15}{4}\right) = 0.2285\,. \end{aligned}$$

On the other hand, under H_1 the probability of the acceptance region for s_4^2 can be computed by a similar method:

$$P\left(3\frac{s_4^2}{\sigma_0^2} \leqslant 7.81 \,\middle|\, H_1\right) = P\left(3\frac{s_4^2}{\left(\frac{4}{3}\sigma_0\right)^2} \leqslant 4.39 \,\middle|\, H_1\right) = 0.77 .$$

The probability of the error of second kind is now the product of these two probabilities: $0.2285 \cdot 0.77 = 0.176$. This means that if the mentioned irregularity occurs in the course of production, then, on the average, our test procedure ends with a wrong decision in one out of six cases. If irregularities occur fairly seldom, then such a proportion can be tolerated, but if the flaw in the machine occurs frequently, then we have to diminish the probability of error of the second kind. In our example this may be performed by using samples of size 9 instead of 4. The calculation, carried out in the same way as before, leads, for the error of second kind, to the probability $0.0234 = 2.34\%$. This value can be taken as a very convenient one for a wide field of practical applications.

In constructing an appropriate test procedure we did not strictly follow the theory described earlier. We started from the practical situation by choosing the sample size and the critical region and settled on the alternative later. In this way we are able to judge whether the test procedure provides the safeguard needed against the real danger. An intermediate way of choosing the test procedure is by means of statistical decision functions (see Section 4.5).

4.4 ELEMENTS OF CORRELATION AND REGRESSION ANALYSIS

This very broad branch of probability theory and mathematical statistics plays a significant role in almost all fields of industrial organization and scientific research work, and also in industrial quality control. Our present aim is only to outline the most basic concepts of this theory and to emphasize its applicability to different questions arising in quality control. For a more detailed treatment of the theory we refer to the books by Ezekiel–Fox [A 15], Hald [A 24] and T. W. Anderson [A 1].

We have already touched upon the question of independence and dependence of random variables. Correlation and regression analysis deals with the qualitative and quantitative study of the stochastic relationship between random variables. The main tool of this field is the well-known method of least squares; and an essential part is the statistical investigation of the results obtained by this method.

Lacking an exact functional relationship between quantities but having a stochastic connection between them, we are in a position to draw inferences about certain quantities from other quantities. It is important that the total production depends on different index numbers relating to the workshop; the quality depends on the raw material, on the production techniques and, last but not least, on the quality control method used. The life span of a part depends on different geometrical, physical, chemical and technical data. It often occurs that parts can be fitted together within relatively large tolerance limits; the life span and safety of the machine depend to a high degree, however, on the deviation of the measures from the prescribed ones.

In such cases the first task is to investigate the existence of independence or dependence. An often used tool in dealing with this problem is the correlation coefficient. If dependence is inferred taken as a next step the type of the stochastic relationship has to be investigated.

4.4.1 Estimation of the Correlation Coefficient

We have introduced the correlation coefficient of the random variables X and Y (Section 2.4.2):

$$\varrho(X, Y) = \frac{E\left[(X - E(x))\,(Y - E(Y))\right]}{D(X)\,D(Y)}$$

We have pointed out that $\varrho(X, Y) = 0$ means the variables are uncorrelated; if, in addition, X and Y are jointly normally distributed then they are also independent.

In practical situations we have first to make a decision between independence and dependence. Having a sample of size n with elements $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ on the random vector (X, Y), it is best to plot the sample elements as points in an (x, y) coordinate system. In many cases even this simple graphical procedure indicates a certain dependence between the two variables, although a functional relationship may not exist between them.

I.e., the value of the one variable does not uniquely determine the other one, but the increase of the one variable may induce an increase (or decrease) in the other variable. In cases of dependence, the next step consists in the determination of the type of the "stochastic" connection between the two variables. The meaning of a "stochastic" connection and how to bring it into a mathematical form will be the topic of the next section.

The independence can be examined with the aid of the χ^2 test. If this procedure indicates dependence, then as a measure of dependence the correlation coefficient $\varrho(X, Y)$ can be used. It can be estimated by the following statistic, the so-called sample correlation coefficient:

$$\varrho(X, Y) \sim r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{s_X s_Y}$$

where the usual notation is applied:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i\,,$$

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

The distribution of the random variable r is known only for random variables with a *bivariate normal distribution* as joint distribution. In this case it can be shown that, if independence holds, i.e., if $\varrho(X, Y) = 0$, the random variable

$$t = \sqrt{n-2}\,\frac{r}{\sqrt{1-r^2}}$$

is distributed according to the Student distribution with parameter $n - 2$. Consequently, the statistic t can be used for testing the independence of the random variables X and Y. Given the level $1 - \alpha$, if we find that the observed value of t lies between $-t_\alpha$ and t_α in the $(n - 2)$nd row of Table VIII, then this does not give rise to any objection against independence. In the opposite case, the deviation from independence can be considered significant.

If X and Y are jointly normally distributed random variables or if n is large enough, under fairly general conditions the following is true of the expected value and variance of r:

$$E(r) = \varrho(X, Y) + 0\left(\frac{1}{n}\right), \quad D^2(r) = \frac{1}{n}(1 + \varrho^2(X, Y))^2 + 0\left(\frac{1}{n^{3/2}}\right).$$

Example: Let X be the breaking strength and Y the tensile stretching of a given type of yarn. Table 7 gives the value of a sample of size 30 on (X, Y) and the method for computing r. Multiplying numerator and denominator by $n - 1 = 29$, we obtain for the empirical correlation coefficient

$$r = \frac{473.2}{\sqrt{25\,557 \cdot 15.44}} = 0.75$$

TABLE 7

i	X_i	Y_i	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i - \overline{X})^2$	$(Y_i - \overline{Y})^2$	$(X_i - \overline{X})(Y_i - \overline{Y})$
1	240	5.4	5	—0.4	25	0.16	—2.0
2	222	5.3	—13	—0.5	169	0.25	6.5
3	203	5.1	—32	—0.7	1 024	0.49	22.3
4	203	4.9	—33	—0.9	1 089	0.81	22.7
5	197	5.8	—38	0.0	1 444	0.00	0.0
6	255	6.3	20	0.5	400	0.25	10.0
7	210	4.6	—25	—1.2	625	1.44	30.0
8	223	6.0	—12	0.2	144	0.04	2.4
9	258	6.6	23	0.8	529	0.64	18.4
10	250	5.9	15	0.1	225	0.01	1.5
11	173	4.3	—62	—1.5	3 844	2.25	93.0
12	260	6.9	25	1.1	225	1.21	27.5
13	233	6.7	—2	0.2	4	0.04	—0.4
14	210	6.0	—25	0.2	625	0.04	—12.5
15	254	6.2	39	0.4	1 521	0.16	15.6
16	220	6.0	—15	0.2	225	0.04	—7.5
17	287	6.9	52	1.1	2 704	1.21	57.2
18	265	7.2	30	1.4	900	1.96	42.0
19	237	5.4	2	—0.4	4	0.16	—0.8
20	243	5.6	8	—0.2	64	0.04	—1.6
21	260	6.2	25	0.4	625	0.16	10.0
22	280	6.4	45	0.6	2 025	0.36	27.0
23	280	6.7	45	0.9	2 025	0.81	40.5
24	240	5.2	—5	—0.6	25	0.36	3.0
25	230	5.7	—5	—0.1	25	0.01	0.5
26	215	5.1	—20	0.7	400	0.49	—14.0
27	271	6.2	6	0.4	36	0.16	2.4
28	256	6.1	21	0.3	441	0.09	6.3
29	221	5.2	—14	—0.6	196	0.36	8.4
30	298	7.0	63	1.2	3 969	1.44	75.6

$\overline{X} = 239.77$ $\overline{Y} = 5.89$ Totals 25 557 15.44 473.2

Considering the samples as coming from a two-dimensional normal distribution we compute the value of t:

$$t = \sqrt{n-2}\,\frac{r}{\sqrt{1-r^2}} = \sqrt{28}\,\frac{0.75}{\sqrt{1-0.5625}} = 6.031$$

In Table VIII (p. 397) the column for the largest occurring level $1 - \alpha = 0.999$ contains, in the 28th row, the critical value $t_{0.001} = 3.674$.

Because our statistic is much larger than the critical value even on this high level, the independence of the two random variables must be rejected; thus, their dependence must be regarded as statistically verified.

4.4.2 The Regression Curve. Regression Lines and Parabolas

Having once verified dependence between two random variables, we may wish to know something about the type and form of their relationship. We do not now assume the normality of the vector (X, Y) but wish to treat the question in general. Denoting the joint density function of X and Y by $f(x, y)$, we may compute the conditional expected values; the conditional expected value of Y given $X = x$ is

$$y = A(x) = E(Y \mid X = x) = \frac{\int_{-\infty}^{\infty} v\, f(x, v)\, dv}{\int_{-\infty}^{\infty} f(x, v)\, dv},$$

while the conditional expected value of X given $Y = y$ is

$$x = B(y) = E(X \mid Y = y) = \frac{\int_{-\infty}^{\infty} u\, f(u, y)\, du}{\int_{-\infty}^{\infty} f(u, y)\, du};$$

these are the so-called regression curves: the first one is the regression of Y on X, the second one is the regression of X on Y.

If we make an observation on X and find it to be $X = x$, then $A(x)$ is the best approximation for the unobserved value of Y belonging to $X = x$ in the following sense: The value of Y for given $X = x$ will fluctuate around $A(x)$. In a long sequence of observations on X our estimates $Y \sim A(x)$ will sometimes be larger or smaller than the actual value of Y, but these deviations will balance one another.

In general, however, we do not know the density function $f(x, y)$; in other cases the expression of $A(x)$ is too complicated, even sometimes containing unknown parameters. In practical applications the usual procedure is to replace the true regression curve by a straight line or by that of a polynomial of a certain degree. These are called regression lines and regression polynomials, respectively:

$$y = A_s(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_s x^s$$

or

$$x = B_r(y) = b_0 + b_1 y + b_2 y^2 + \ldots + b_r y^r.$$

The coefficients $\{a_i\}$ or $\{b_j\}$ will be estimated from the sample using the method of least squares.

The first question which arises is the convenient choice of the degree of the polynomial. To answer this we plot the sample elements as points in an (x, y) coordinate system, as already mentioned. The pattern of these points will indicate whether we should use a linear regression or a polynomial of second or third degree, etc.

The application of the least squares method leads to a system of linear equations in the unknown coefficients. We will illustrate this with our example, where we have chosen linear regressions in both directions. Having the sample elements (X_i, Y_i) $i = 1, 2, \ldots, n$, we determine the coefficients a_i in such a way that the sum of squares of the errors will be minimized:

$$\sum_{i=1}^{n} (Y_i - a_0 - a_1 X_i)^2 = \text{minimum}.$$

Taking the derivatives with respect to a_0 and a_1, a slight modification leads to the following system of equations for the estimators $\hat{a}_0$ and $\hat{a}_1$ of the coefficients:

$$\hat{a}_0 + \frac{1}{n} \sum_{i=1}^{n} X_i \hat{a}_1 = \frac{1}{n} \sum_{i=1}^{n} Y_i ,$$

$$\frac{1}{n} \sum_{i=1}^{n} X_i \hat{a}_0 + \frac{1}{n} \sum_{i=1}^{n} X_i^2 \hat{a}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i$$

or

$$\hat{a}_0 + \overline{X}\hat{a}_1 = \overline{Y}$$

$$X\hat{a}_1 + \overline{X^2}\hat{a}_1 = \overline{X}\,\overline{Y}$$

with the notations

$$\overline{X^2} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \text{ and } \overline{XY} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i ,$$

which are the second and mixed empirical (sample) moments respectively. Using our earlier notation for the unknown coefficients we obtain

$$\hat{a}_1 = \frac{\overline{XY} - \overline{X}\ \overline{Y}}{\overline{X^2} - \overline{X^2}} = \frac{\sum_{i=1}^{u} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = r \frac{s_y}{s_x} .$$

$$\hat{a}_0 = \overline{Y} - \hat{a}_1 \overline{X} = \overline{Y} - r \frac{s_y}{s_x} \overline{X}.$$

The equation of the regression line takes on the following simplified form:

$$y = \overline{Y} + r\frac{s_y}{s_x}(x - \overline{X})$$

or the standardized form:

$$\frac{y - \overline{Y}}{s_y} = r\frac{x - \overline{X}}{s_x}.$$

It can be observed that the regression line passes through the mass-center $(\overline{X}, \overline{Y})$ of the point set and has the slope

$$R(Y \mid X) = r\frac{s_y}{s_x}.$$

which is called the empirical regression coefficient of Y on X.

In a similar way we get the regression line of X on Y as

$$x = \overline{X} + r\frac{s_x}{s_y}(y - \overline{Y})$$

or

$$\frac{x - \overline{X}}{s_x} = r\frac{y - \overline{Y}}{s_y}.$$

This intersects the regression line of Y on X at the point $(\overline{X}, \overline{Y})$ and has the slope

$$R(X \mid Y) = r\frac{s_x}{s_y}.$$

Since Table 7 contains all data necessary for the determination of the regression lines we can turn to the numerical values. We get

$$y = 0.0185(x - 239.77) + 5.89$$

for the regression of Y on X and

$$x = 30.65(y - 5.89) + 239.77$$

for the regression of X on Y.

The data of Table 7, together with both regression lines, are represented in Fig. 31.

Similarly, we can compute the coefficients $\hat{a}_2, \hat{a}_3, \ldots$ of the polynomials if a higher degree is used; in each case the number of equations will be equal to the number of unknown coefficients. No principal difficulties arise here, but the numerical evaluation in the case of high degree requires a computer.

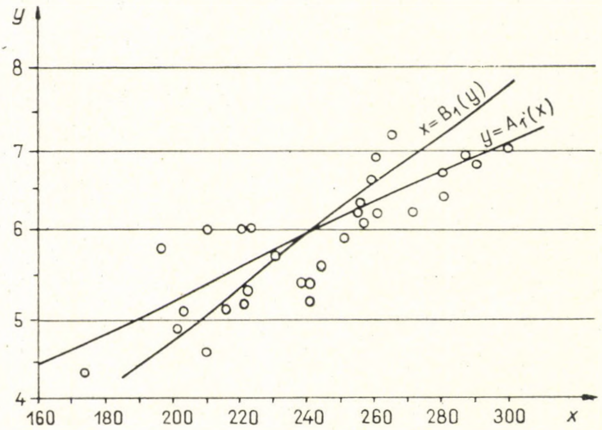


FIG. 31. Sample of size 30 from a two-variate distribution (X = breaking strength, Y = tensile stretching of a given kind of yarn; see Table 7) with corresponding regression lines (see Section 4.4.2)

4.5 ELEMENTS OF THE THEORY OF DECISION FUNCTIONS

4.5.1 The Decision Problem

As we have seen, each statistical procedure ends with a certain *decision*. These decisions are based on samples; therefore they are *statistical* decisions. In estimation problems, the decision concerns the acceptance of a given value instead of the true parameter. Acceptance or rejection of a hypothesis is a typical decision between two alternatives. Determining a confidence interval is the decision to consider the true value of the parameter as lying within this interval.

In each case a function of the sample elements, i.e., a statistic, is the tool for our decision. How to choose an appropriate statistic is a basic problem of mathematical statistics, and in many cases we have considered (or at least touched upon) this question. At the same time, the problem was not answered completely; only certain points of view concerning the choice of the appropriate statistic were mentioned. For example, in the case of the estimation of the variance of a random variable, the sample variance is an unbiased, strongly consistent estimator, but by using the sample range the estimation is much easier and needs fewer calculations. In the case of hypothesis testing, small errors of first and second kind were desired. At the same time we pointed out that constructing a test procedure with small probabilities for both kinds of errors will result in a test with a fairly high, even with an unacceptably high, sample size.

In such situations, at least in industrial and economical applications, the financial or economic consequences must be considered in choosing the statistical procedure. If we are able to numerically express the financial (or other) effects of wrong decisions, and if a favorable choice of the statistical procedure can be made by reducing this quantity as much as possible, then we have a good means for determining the appropriate statistical procedure or for comparing two procedures.

At the same time, this is the way to get a general and unified treatment of the different kinds of problems arising in mathematical statistics. The initiator of this theory was A. Wald, who — in his book "Statistical Decision Functions" [A 49] — gave the first detailed development of this discipline. For an up-to-date treatment of the theory we refer to the book of Ferguson [A 16].

Even in cases when it is complicated or even impossible to master the problem completely with the above techniques, it may prove to be very useful to follow along these lines as far as possible. In other cases we may be able to formulate the problem, but the solution may bring about mathematical difficulties.

In the next sections, only some of the basic concepts of this theory will be given. The knowledge of the basic idea and the principles of this field may be very helpful in many cases when decisions based on statistical data are needed.

4.5.2 Statistical Decision and the Decision Function

Our aim now is to introduce the following simple, but slightly abstract, common formulation of all statistical procedures: Having the sample space $\mathfrak{X}$, we observe an element X of it and thereupon select a decision d from the set D of all possible decisions which can be made concerning the practical problem. D is called the *decision space*.

The choice of d has to be made according to a certain rule; this rule is called the statistical decision function and will be denoted by $d(X)$: it will select that element d of D which belongs to the outcome X of the observation.

Let us turn now to a more detailed treatment.

Suppose that we have to estimate the expected value μ of the random variable X distributed according to $N(\mu, \sigma_0)$, where σ_0 is known. Suppose further that we have n observations on X, i.e., the sample space is

$\mathfrak{X} = R_n$, the Euclidean n space with elements $X^{(n)} = (X_1, X_2, \ldots, X_n)$. By means of the observed vector $X^{(n)}$ we will estimate μ with the aid of a certain function of $X^{(n)}$. Such statistics may be, e.g.

$$d_0(X^{(n)}) = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) = \bar{X}$$

or

$$d_1(X^{(n)}) = \sum_{i=1}^{n} p_i X_i \text{ with } \sum_{i=1}^{n} p_i = 1$$

or

$$d_2)X^{(n)}) = \frac{1}{2}(X_1^* + X_n^*).$$

At the same time, we may observe that each actual (or possible) decision d is a real number; accordingly, the decision space coincides with the set of real numbers: $D = R_1 = \{d: -\infty < d < \infty\}$. But this is also the set of the possible values of the parameter μ. In general, it is true that the decision space is closely connected with the parameter space, especially with the assumed hypotheses.

We remark here that the error made by using this procedure is $d_0(X^{(n)}) - \mu$ or $d_1(X^{(n)}) - \mu$ or $d_2(X^{(n)}) - \mu$, if we use the corresponding decision function and if μ is the true value of the parameter.

Let us turn now to the test procedure. Suppose lots of products are transferred to us from a factory. We have to qualify each lot according to whether or not its fraction defective exceeds the value p_0. This means that we have to make a decision between the two hypotheses

$$H_0\colon p \leqslant p_0$$

and

$$H_1\colon p > p_0.$$

The choice will be made with the aid of a sample.

We wish to apply a test procedure of the following type: If we find that the number $\varkappa_n$ of defective items in a sample of size n (n unspecified !) is smaller than a certain constant $c(n)$, then we accept H_0, in the opposite case we reject it. The first case will mean the acceptance, the latter case the rejection of the lot. The sample space is now $\mathfrak{X} = \{0, 1, 2, \ldots, n\}$. The decision space is a set containing two elements $D = \{d_0, d_1\}$; here d_0 denotes the decision of accepting H_0, while d_1 the decision of accepting H_1. The decision function based on the statistic has the form

$$d(\varkappa_n) = \begin{cases} d_0, \text{ if } \varkappa_n = 0, 1, 2, \ldots, c(n), \\ d_1, \text{ if } \varkappa_n = c(n) + 1,\ c(n) + 2, \ldots, r. \end{cases}$$

The effect of accepting a bad lot containing a fraction defective p greater than p_0 will in general be proportional to $p - p_0$.

We emphasize here that n was unspecified; the choice of n is left open. Our aim is to determine the decision procedure, including choosing the sample size, as economically as possible.

4.5.3 Loss Function and Risk Function

As pointed out in our examples, statistical problems concern a parameter, say ϑ, the true value of which is unknown: the statistical procedure ends with a decision, say d. In many cases we are in the position to numerically determine the loss caused by making a bad decision. Let us denote by $L(\vartheta, d)$ the amount of the loss when the true value of the parameter is ϑ and the decision d was made; $L(\vartheta, d)$ is called the *loss function*. Our decision is based on an observation X which is an element of the sample space $\mathfrak{X}$. Let $F(x)$ be the distribution function of the (real or vector) variable X. Let us suppose, furthermore, that we use the decision function $d(X)$. Then $L(\vartheta, d(x))$ gives the loss, when ϑ is the true value of the parameter and the observation resulted in $X = x$.

It is fairly difficult to form an opinion as to the effectiveness of the decision function $d(X)$ by means of the loss function: it requires the examination of $L(\vartheta, d(X))$ for each possible outcome $X \in \mathfrak{X}$ and given ϑ. But it is usually the case in statistical problems that we are interested only in the consequences *on the average*. This means that the expected value of the loss is to be considered; this gives us the mean value of the loss in the long run when the decision function $d(X)$ is used.

The average loss is called the *risk function*; expressed as a function of the true parameter value and of the decision function used, it has the form:

$$R_{d(X)}(\vartheta) = E_\vartheta[L(\vartheta, d(X))] = \int_{\mathfrak{X}} L(\vartheta, d(x))dF(x;\vartheta).$$

Knowing the loss function we can determine the risk functions belonging to the different possible decision functions which come into play in the actual problem. This allows the possibility of comparing two decision functions: for a given value ϑ the decision function $d_0(X)$ is better than the decision function $d_1(X)$ if the risk belonging to $d_0(X)$ is smaller than the risk when using $d_1(X)$, i.e., if

$$R_{d_0(X)}(\vartheta) < R_{d_1(X)}(\vartheta).$$

The decision function $d_0(X)$ is called *uniformly better* than the decision function $d_1(X)$ if the following relation holds for each value of the parameter ϑ:

$$R_{d_0(X)}(\vartheta) \leqslant R_{d_1(X)}(\vartheta) \tag{30}$$

and if at least one parameter value, say ϑ_0 ,exists for which the inequality sign holds:

$$R_{d_0(X)}(\vartheta_0) < R_{d_1(X)}(\vartheta_0)\,. \tag{31}$$

If $R_{d_0(X)}(\vartheta) = R_{d_1(X)}(\vartheta)$ for each value of ϑ then we say that the decision functions $d_0(X)$ and $d_1(X)$ are equivalent.

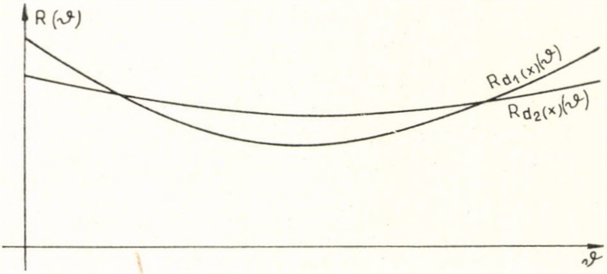


FIG. 32. Risk functions belonging to admissible decision functions

If there exists a decision function $R_{d_0(x)}$ (ϑ) for which relations (30) and (31) are satisfied, no matter what the decision function $d_1(X)$ may be, then we say that $d_0(X)$ is a *uniformly best decision function.*

If in a given problem we find that the decision function $d_0(X)$ is uniformly better than $d_1(X)$, it is obvious that the use of $d_1(X)$ is out of the question. A decision function like $d_1(X)$ is *inadmissible.* A decision function $d(X)$ is *admissible* if there does not exist a decision function which is uniformly better than $d(X)$.

When a uniformly best decision function does not exist there is nothing better for us to do but apply an admissible decision function. Owing to the nature of admissible decision functions, however, many difficulties may arise. If both $d_1(X)$ and $d_2(X)$ are admissible decision functions, then this implies that $d_1(X)$ is better for certain values of ϑ, while for other ϑ's $d_2(X)$ is better. (Figure 32.) How do we choose between them and select an appropriate decision function?

In certain cases there exists a prior distribution function for the parameter set: Suppose that ϑ is real; it varies from $-\infty$ to ∞. From earlier experiences we know that $g(\vartheta)\Delta\vartheta$ is the probability that the true value of the parameter lies in the small interval $\left(\vartheta - \frac{1}{2}\Delta\vartheta,\ \vartheta + \frac{1}{2}\Delta\vartheta\right)$, i.e., $g(\vartheta)$ is the prior density function.

When applying the decision function $d(X)$ the mean value of the risk is:

$$\bar{R}_{d(X)} = \int_{-\infty}^{\infty} R_{d(X)}(\vartheta)g(\vartheta)d\vartheta = \int_{-\infty}^{\infty} \int_{X} L(\vartheta, d(x))\, dF(x, \vartheta)\, dG(\vartheta)$$

with $G(\vartheta) = \int_{-\infty}^{\vartheta} g(t)dt$. The decision function $d(X)$ which minimizes this average risk is called the *Bayes solution* belonging to the prior distribution $G(\vartheta)$:

$$\bar{R}_{d(X)} \leqslant \bar{R}_{d^*(X)}$$

for any decision function $d^*(X)$. If such a $d(X)$ does not exist we look for a decision function with an average risk near to the infimum. In the absence of a prior distribution, the question is more complicated.

4.5.4 Examples of Loss and Risk Functions. The Completed Risk Function

Before continuing our general considerations we will illustrate these notions.

Let us begin with the *estimation problem*. Suppose that the real parameter ϑ $(-\infty < \vartheta < \infty)$ is to be estimated. The decision space is now the real line: $D = \{d\colon -\infty < d < \infty\}$. The difference $|d - \vartheta|$ can be considered as the error if the decision d is made and ϑ is the true value of the parameter. For various reasons it is often common and convenient practice to use the square of this expression as loss function:

$$L(\vartheta, d) = (d - \vartheta)^2.$$

If we now consider the decision function $d(X)$ as estimator, then we obtain for the risk function

$$R_{d(X)}(\vartheta) = \int_{\mathfrak{X}} (d(x) - \vartheta)^2\, dF(x; \vartheta)$$

where $F(x; \vartheta)$ denotes the distribution function of the random variable X. This means that if the estimator $d(X)$ is unbiased, then the risk function is the variance of the estimator. In certain cases, however, the absolute value of the deviation is chosen as loss function:

$$L^*(\vartheta, d) = |d - \vartheta|$$

and we have for the risk function

$$R^*_{d(X)}(\vartheta) = \int_{\mathfrak{X}} | d(x) - \vartheta | \, dF(x; \vartheta).$$

We turn now to the problem of testing statistical hypotheses. Consider two hypotheses for the real parameter ϑ: the null hypothesis

$$H_0\colon \vartheta \leqslant \vartheta_0$$

and the alternative

$$H_1\colon \vartheta > \vartheta_0.$$

After observing the random variable X, either of the decisions d_0 or d_1 is possible, i.e., the acceptance of H_0 or H_1, respectively. Thus the decision space is $D = \{d_0, d_1\}$. The distribution function of X will be denoted by $F(x; \vartheta)$. For making decisions between H_0 and H_1, the decision function $d(X)$ will be used; the decision function is determined by a test statistic $t(X)$, e.g., in the following way:

$$d(X) = \begin{cases} d_0, \text{ if } t(X) \geqslant c \\ d_1, \text{ if } t(X) < c. \end{cases}$$

This means that we have the critical region

$$\mathfrak{X}_c = \{X\colon t(X) \geqslant c\}$$

and the acceptance region

$$\mathfrak{X}_a = \{X\colon t(X) < c\}.$$

As we know, two kinds of bad decisions can be made: errors of first and second kind. In mathematical statistics the loss function is usually constructed so that it is zero if a good decision is made.

This practice is debatable, but at least it takes into account the worst situation from the point of view of the receiver (consumer).

In testing problems the so-called *simple loss function* is often used. This has the following form for the above problem:

$$L(\vartheta, d_0) = \begin{cases} 0, & \text{if } \vartheta \leqslant \vartheta_0, \\ a_0, & \text{if } \vartheta > \vartheta_0, \end{cases}$$

$$L(\vartheta, d_1) = \begin{cases} a_1, & \text{if } \vartheta \leqslant \vartheta_0, \\ 0, & \text{if } \vartheta > \vartheta_0. \end{cases}$$

This means that the simple loss function does not measure varying economic consequences for different values of ϑ if a bad decision is made. Certain practical situations correspond to this model, but this is not always the case. This is why we now turn to the more general case, giving the following very important example introduced in Section 4.5.2.

Lots of products, each consisting of N pieces, are to be accepted. In accordance with requirements for the quality of the lot, the null hypothesis is

$$H_0\colon p \leqslant p_0,$$

and the alternative hypothesis is

$$H_1\colon p > p_0$$

where p $(0 \leqslant p \leqslant 1)$ is the fraction defective in the lot.

In constructing the loss function we must take into account the fact that if we accept a bad lot, the bad pieces exceeding the acceptable number of bad pieces Np_0 will cause us damage. Let us denote by a the price of a single product. In rejecting a good lot we also have certain damages, e.g., delay in production or in selling, superfluous activity, etc; this kind of loss will be given a fixed value, say A. Considering these economic assumptions the loss function will have the following form:

$$L(p, d_0) = \begin{cases} 0, & \text{if } p \leqslant p_0, \\ N(p - p_0)a, & \text{if } p > p_0, \end{cases}$$

$$L(p, d_1) = \begin{cases} A, & \text{if } p \leqslant p_0, \\ 0, & \text{if } p > p_0. \end{cases}$$

In deciding to accept or reject the lot, the number $\varkappa_n$ of defective items in a sample of size n will be chosen as the test statistic. As a basis for constructing a decision function, we note that it is reasonable to accept H_0 for small values of $\varkappa_n$ and reject it otherwise. This statement requires proof (which can be given), but here we omit it and define the decision function as follows:

$$d(\varkappa_n) = \begin{cases} d_0, \text{ if } \varkappa_n < c \\ d_1, \text{ if } \varkappa_n \geqslant c. \end{cases}$$

In looking for a "best" decision function, we restrict ourselves to the determination of the sample size n, and to the selection of the critical value c which depends now on $n : c = c(n)$. The means of selection is the risk function which now has the following form:

$$R_{d(\varkappa)}(p) = \begin{cases} A \displaystyle\sum_{i=c(n)}^{n} \binom{n}{i} p^i (1-p)^{n-i}, & \text{if } \quad p \leqslant p_0 \\ Na(p - p_0) \displaystyle\sum_{i=0}^{c(n)-1} \binom{n}{i} p^i (1-p)^{n-i}, & \text{if } \quad p > p_0 . \end{cases}$$

We take a loss in the case of $p \leqslant p_0$ if we reject the lot; this occurs when we find $\varkappa_n \geqslant c(n)$, the probability of which is, according to the binomial law, $\sum_{i=c(n)}^{n} \binom{n}{i} p^i(1-p)^{n-i}$. For the second expression the corresponding reasoning holds. Figure 33 shows the shape of this risk function for certain values of n, c and other constants.

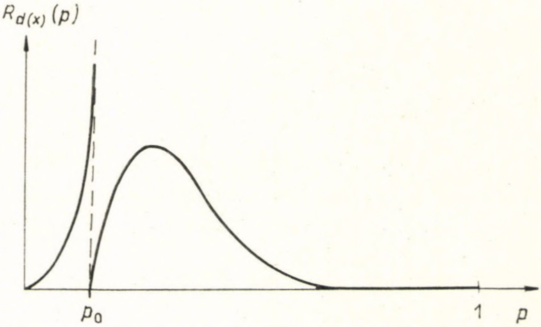


FIG. 33. Risk function

Finally, we note that in quality control problems it may be convenient to include certain costs into the risk function. First of all, the selection and examination of the sample elements usually result in a cost which can be taken to be proportional to the sample size. The work of computing the statistic is not always negligible. As an illustration we consider the following simple estimation problem: let the components of the vector variable $X = (X_1, X_2, \ldots, X_n)$ be independent, identically-distributed random variables with (unknown) expected value μ and known variance σ_0^2. Suppose the parameter is to be estimated, and the loss function

$$L(\vartheta, d) = (d - \vartheta)^2$$

is chosen. If $d(X)$ is an unbiased estimator for ϑ, then, as mentioned, the risk function has the form

$$R_{d(X)}(\vartheta) = D_\vartheta^2(d(X)).$$

Let us assume now that the cost of selecting and examining a single sample element is c, and let us denote the cost caused by the evaluation of the statistic $d(X)$ by $c_n(d(X))$. Then the *completed risk function* will be

$$R_{d(X)}^c(\vartheta) = D_\vartheta^2(d(X)) + cn + c_n(d(X)).$$

If, e.g.

$$d(X) = \frac{X_1 + X_2 + \ldots + X_n}{n} = \bar{X}, \quad \text{set} \quad c_n(\bar{X}) = c'\sqrt{n}$$

and

$$R^c_{\bar{X}}(\vartheta) = R^c_{\bar{X}} = \frac{\sigma_0^2}{n} + cn + c'\sqrt{n}.$$

If not only the expected value of the loss but also some further costs are included in the risk function then we speak about the *completed risk function*.

4.5.5 The Role of the Unbiased Estimator and of the Sufficient Statistic

When first considering a statistical problem, we should take into account, at least in principle, all possible decision functions. This being too large a collection to choose from, we turn to some natural limitations restricting the set of possible decision functions.

In estimation problems, if all possible statistics are taken into account as estimators, then we cannot find a uniformly best one among them.

If the parameter ϑ is to be estimated and ϑ_0 is a given possible value of it then the decision function $d(X) \equiv \vartheta_0$ (i.e. the one which results in the estimate ϑ_0 independently of X), will be a best decision function when the actual value of the parameter is exactly ϑ_0; but it does not provide a good estimator for parameter values differing considerably from ϑ_0. This is why it is quite natural to restrict ourselves to unbiased estimators if such exist.

Let us consider the estimation problem treated at the end of the preceding section. If $d(x)$ is an unbiased estimator and we take into account only the cost cn, then the completed risk function will have the form

$$R^*_{d(X)}(\vartheta) = D^2_\vartheta(d(X)) + cn.$$

Let us assume in addition that the X_i's are distributed according to $N(\vartheta, \sigma_0)$. Then by virtue of the Cramér–Rao inequality (see Section 4.2.2.5) the best unbiased estimator for the expected value is the sample mean $\bar{X}$, i.e.

$$D^2_\vartheta(\bar{X}) = \frac{\sigma_0^2}{n} \leqslant D^2_\vartheta(d(x))$$

for any unbiased estimator $d(X)$ and for any value of the parameter ϑ. This means that for the loss function

$$L(\vartheta, d) = (d - \vartheta)^2$$

and with the completed risk function

$$R^*_{d(X)}(\vartheta) = D^2_\vartheta(d(X)) + cn$$

the decision problem has a uniformly best solution among all unbiased decision functions, and this is the sample mean $d(X) = \overline{X}$.

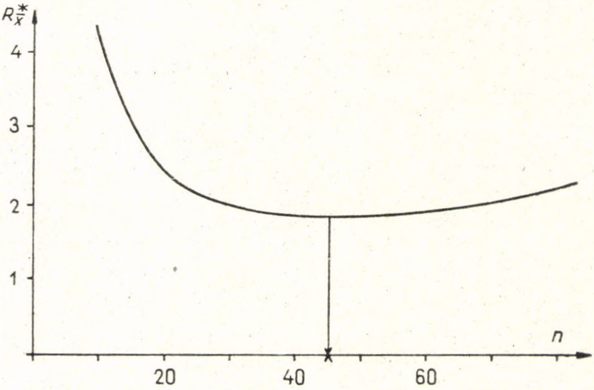


FIG. 34. Determination of the most economical sample size based on the risk function

$$R_X = \frac{\sigma_0^2}{n} + cn + c'\sqrt{n}$$

The minimum risk is now (Fig. 34)

$$R^*_{\overline{X}}(\vartheta) = \frac{\sigma_0^2}{n} + cn.$$

The situation will differ slightly if the completed risk function contains the computational costs as well. In this case the completed risk takes on the form

$$R^{**}_{d(X)}(\vartheta) = D^2_\vartheta(d(X)) + cn + c_n(d(X)).$$

Substituting $d(X) = \overline{X}$ and setting $c_n(\overline{X}) = c'\sqrt{n}$, we obtain

$$R^{**}_{\overline{X}}(\vartheta) = \frac{\sigma_0^2}{n} + cn + c'\sqrt{n}.$$

Although the inequality

$$\frac{\sigma_0^2}{n} \leqslant D_\vartheta^2(d(X))$$

now holds for each ϑ. the computational costs for some $d(X)$ may be much less than $c'\sqrt{n}$, e.g., $c''\log n$, and therefore the completed risk function may be more favorable in the latter case.

Such a situation may occur if a statistic of very simple form is used, e.g., the median, or the midrange $\frac{1}{2}(X_1^* + X_n^*)$ computed from the first and last order statistics. Finally, for certain values of the constants c, c', c'' and at least for certain values of n the inequality

$$R_{d(x)}^{**}(\vartheta) \leqslant R_{\overline{X}}^{**}(\vartheta)$$

may be valid for certain values of ϑ.

This circumstance must be taken into account if, for the estimation of the same parameter, two different statistics require a very different amount of computational work, as is the case for the sample variance and the sample range. And here we turn to the use of sufficient statistics.

Sufficient statistics play a basic role in statistical decision problems. Without giving an exact formulation, the following statement is essentially true: If the loss function $L(\vartheta, d)$ is a convex function of d for each ϑ and if a sufficient statistic $T(X)$ exists for the parameter ϑ, then to each decision function $d(X)$ there corresponds a decision function which is a function of $T(X)$ with risk not larger than that of $d(X)$. This decision function is the conditional expectation of $d(X)$ under the condition $T(X) = t$, i.e., the statistic $f(t) = E(d(X) \mid T(X) = t)$, for which

$$R_{f(t)}(\vartheta) \leqslant R_{d(X)}(\vartheta) \text{ for each } \vartheta.$$

This means that in order to find a good or best decision function we may confine ourselves to functions of the sufficient statistic if such a statistic exists. This is a very important fact to which we will return in the next section.

At the same time we have to emphasize that in using quality control methods, certain costs must often be taken into consideration, i.e., a completed risk function will be used. In such cases, even if the loss function remains convex, the risk function will not possess this property. Consequently, in certain situations, quality control methods will be based on a non-sufficient statistic, although a sufficient statistic exists.

4.5.6 Selection of the Appropriate Decision Function. The Minimax Method

A simple statistical problem concerning the real-valued parameter $\vartheta(-\infty < \vartheta < \infty)$ will be considered. Accordingly, the decision space will be $D = \{d\colon -\infty < d < \infty\}$. We choose a loss function $L(\vartheta, d)$ which is convex in the above sense. The observation $X = (X_1, X_2, \ldots, X_n)$ is a random vector with real-valued components; the distribution function of X will be denoted by $F(x, \vartheta)$ with $x = (x_1, x_2, \ldots, x_n)$, a point of the n-dimensional Euclidean space. If $d(X)$ is a decision function, then the risk function has the form

$$R_{d(X)}(\vartheta) = E_\vartheta[L(\vartheta, d(X))] = \int_X L(\vartheta, d(x))\, dF(x; \vartheta).$$

Throughout this section we do not consider any costs but the loss.

For the selection of an appropriate decision function the following two cases may be considered solved, at least in principle.

(a) There exists a uniformly best decision function, say $d_0(X)$, for which

$$R_{d_0(X)}(\vartheta) \leqslant R_{d(X)}(\vartheta)\,,$$

for any ϑ and for any decision function $d(x)$.

As already mentioned, having a sufficient statistic $T(X)$ we have to look for a best decision function among the functions of $T(X)$. Lacking a sufficient statistic, an optimal solution rarely exists in this sense.

(b) In some practical situations we may have a prior distribution. This can be determined by heuristic considerations or from earlier experiences; the later observations will complete or improve the initial assumption. Denoting by $G(t)$ the prior distribution function of ϑ, and applying the decision function $d(X)$, the average risk will be

$$\bar{R}_{d(X)} = \int_{-\infty}^{\infty} R_{d(X)}(t)\, d\,G(t)\,.$$

The decision function $d_0(X)$ will be optimal if it minimizes this quantity, or at least, practically, if it is a value for which the average risk is near the infimum

$$\inf_{d(X)} \bar{R}_{d(X)}\,,$$

when $d(x)$ runs over all decision functions which come into account.

(c) Let us now consider the case when neither (a), nor (b) is satisfied. As we have seen before, it is worthwhile to consider admissible decision functions. A set C of decision functions is called *complete* if to any decision function which is not contained in C there exists an element of C which is better. We say that the set C_0 of decision functions is a *minimal complete class of decision functions* if C_0 is complete but it does not contain any complete subset. As it can be easily verified, if a minimal complete class exists, then it agrees with the set of admissible decision functions.

In considering a statistical decision problem, our main aim is to determine the minimal complete class of decision functions. Having a set of decision functions, it is not always simple to prove the minimality; in certain cases the completeness is easier to handle.

Having a complete class of decision functions containing more than one element, the problem of selecting an appropriate decision function does not have a unique solution. The selection may be based on different principles, each of which offers a defense against certain disadvantageous situations; this means that each selection is optimal in a certain respect, but none of them has an overall optimality property. We will now outline the often used *minimax* principle, which has a motivation stemming from the game theory. For a further development of this theory we refer again to the book [A 16].

Briefly: we look for the decision rule which minimizes the maximum risk. If we use the decision function $d(X)$ then the maximum risk will be

$$\max_{(\vartheta)} R_{d(X)}(\vartheta).$$

Now, for each decision function of the complete class, we can determine this value, and hence we may select the decision function $d_0(X)$ (if such exists) for which this value takes on its minimum:

$$\max_{(\vartheta)} R_{d_0(X)}(\vartheta) = \min_{d(X)} \max_{(\vartheta)} R_{d(X)}(\vartheta).$$

Using such a $d_0(X)$ we may be sure that in the worst case our average loss will not exceed this value.

Our aim has been to introduce the reader to a way of thinking which may be useful in practical work and in formulating problems of quality control from an economical point of view.

4.6 ON STOCHASTIC PROCESSES

4.6.1 Introduction

In taking a sample from a lot or from the items produced by a machine our main requirement up to now has been to ensure the independence of the sample elements since most of the routine statistical procedures are based on independence. Actually, this assumption is only approximately satisfied in many practical cases. A certain degree of dependence exists between the random quantities of a sample. This dependence is often negligible, and many statistical procedures are not very sensitive to a slight degree of dependence. If the existence of a significant degree of dependence is suspected, then it can be tested by various procedures. In case these indicate that dependence is present, there are two possibilities for dealing with the practical problem: we may either take steps to eliminate this circumstance; or we may handle the sequence of random variables as dependent elements, i.e., consider it as a *stochastic process.*

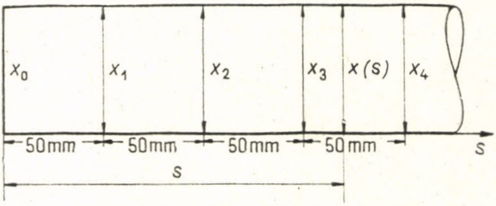


FIG. 35. Thickness of steel plate at equidistant points and at point s

Moreover, in many practical problems, including many situations in industrial quality control, sets or sequences of random quantities occur in which dependence is a basic feature. That is why the theory of stochastic processes has played a growing role in this field, although its methods are still not widespread or well-elaborated.

We turn now to some examples. (Figure 35.) Let us consider the thickness of a narrow, but long, steal plate. If we measure the thickness in the longitudinal direction in distances of 50 mm each, then we get a sequence of random variables

$$X_1, X_2, \ldots, X_n, \ldots \tag{32}$$

In many cases the random variables obtained in this way can be considered independent and are suitable for examining the homogeneity of the thickness, the determination of their mean value, etc., by means of methods elaborated for independent random variables.

Suppose now that the distance between two points is much less than 50 mm, e.g., 2—3 mm. In this case we obtain a sequence or random variables denoted by

$$X_1^{(0)}, X_2^{(0)}, \ldots X_n^{(0)}, \ldots \tag{33}$$

Now the random variables $X_n^{(0)}$ cannot, in general, be taken as independent; at least, neighboring values or values belonging to points lying close to each other must be considered as dependent.

Obviously, significant dependence may take place in the case of sequence (32) as well if, e.g., the thickness shows a certain periodic behavior, but let us ignore this aspect now and consider (32) as an independent sequence.

Finally, the thickness can be examined at each point: let us denote by s the distance of a point of the plate from the origin (see Figure 35) and by $X(s)$ the thickness at this point. Then we have a continuous function defined on the interval $(0, L)$ where L denotes the length of the plate

$$X(s),\ 0 \leqslant s \leqslant L. \tag{34}$$

In (32), (33) and (34) we have represented stochastic processes: they are sets of random variables, the first two being *sequences* of random variables, while the third one is a *continuous* function (this means that to each value of the parameter s $(0 \leqslant s \leqslant L)$ there belongs a random variable $X(s)$). The first example (32) is the simplest: according to our assumption it contains independent random variables. In a stochastic process the set of random variables is parameterized; in the first two examples the parameters are discrete: $n = 1, 2, \ldots$, i.e. they take on integer values only. In the last example the parameter as a distance takes on each real value in an interval; therefore it is a continuous parameter.

The possible values of the occurring random variables are real numbers, in each case we have continuous random variables. We turn now to a more complete description and classification of stochastic processes.

4.6.2 The Concept of a Stochastic Process

As already indicated, a stochastic process with a discrete parameter is a sequence of random variables:

$$X_1, X_2, \ldots, X_n, \ldots$$

The X_i's may take on discrete (e.g. integer) values, in which case we speak of a *discrete stochastic process with discrete parameter*. Or the X_i's may take on any real number value; in this case we have a *continuous stochastic process with discrete parameter*.

A stochastic process with continuous parameter is usually written

$$X_t,\ 0 \leqslant t \leqslant T.$$

Depending on the possible values of X_t it may be that we have a *discrete stochastic process with continuous parameter.* The parameter t represents, in many cases, time or length. For example, X_t may denote the number of yarn breaks in a spinning mill on a machine from an initial time $t = 0$ until the instant t. Then, from the beginning of the shift at $t = 0$, X_t is zero until the first breaking; if this happens at time t_1 then $X_t = 1$, if $t_1 \leqslant t < t_2$, where t_2 indicates the time of the second breaking, etc. We have seen an example (34) of a *continuous process with continuous parameter.*

The behavior of a single random variable X is known if we know its distribution function $P(X \leqslant x) = F(x)$. For a finite set $(X_1, X_2, \ldots, X_n)$ of random variables we need in general, knowledge of the joint distribution

$$P(X_1 \leqslant x_1, X_2 \leqslant x_2, \ldots X_n \leqslant x_n) = F(x_1, x_2, \ldots, x_n).$$

A stochastic process can be considered as completely determined if, for each finite set $t_1, t_2, \ldots, t_n$ $(n = 1, 2, \ldots)$, the joint distribution of the corresponding random variables

$$P(X_{t_1} \leqslant x_1, X_{t_2} \leqslant x_2, \ldots X_{t_n} \leqslant x_n) = F_{t_1,t_2 \ldots t_n}(x_1, x_{2, \ldots} x_n)$$

is known.

The values $t_1, t_2, \ldots, t_n$ may be taken from a discrete set or from a (finite or infinite) interval $(0, T)$ of the real axis, depending on the nature of the process.

Let us suppose now that we have a time process (e.g., the concentration of a chemical substance in an ingredient used in the course of the production process) and that the value of the random variable has been observed at the moments $t_1 < t_2 < t_3 < \ldots < t_{n-1}$. In order to make an inference on the behavior of the process at a later instant $t_n (> t_{n-1})$, the knowledge of the following conditional probability is needed:

$$P(X_{t_n} \leqslant x \mid X_{t_1} = x_{t_1}, X_{t_2} = x_{t_2}, \ldots X_{t_{n-1}} = x_{t_{n-1}}),$$

where $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ are the results of the earlier observations. In certain cases it turns out that this probability depends on the last observed value $X_{t_{n-1}} = x_{t_{n-1}}$ only, and not on earlier observations:

$$P(X_{t_n} < x \mid X_{t_1} = x_{t_1}, X_{t_2} = x_{t_2}, \ldots X_{t_{n-1}} = x_{t_{n-1}}) = P(X_{t_n} < x \mid X_{t_{n-1}} = x_{t_{n-1}}).$$

In this case we speak of a *Markov process,* or if the parameter is discrete, of a *Markov chain.* This means that if in a Markov process or in a Markov chain the value of the random variable is given for a parameter value (a time), then the future behavior of the process depends on the past only through this value. This very important type of random processes will be discussed.

Another important type of random process is the so-called *stationary stochastic process*. A stochastic process (continuous or discrete) is strictly stationary if the joint distribution is invariant under the translation of the parameter

$$P(X_{t_1} \leqslant x_1, X_{t_2} \leqslant x_2, \ldots X_{t_n} \leqslant x_n) = P(X_{t_1+s} \leqslant x_1, X_{t_2+s} \leqslant x_2, \ldots X_{t_n+s} \leqslant x_n)$$

for any nonnegative value of s. From this it follows that the distribution of a single variable is the same for all possible values of the parameter: $P(X_t \leqslant x) = F(x)$, independently of t.

The stochastic process given in (32), i.e., the independent sequence, is a Markov-chain and stationary process at the same time.

Now we turn to some important theoretical quantities connected with stochastic processes.

In the case of a single random variable, the expected value and the variance are usually the most important characteristics making it possible to predict the prospective behavior of the random variable (e.g., with the aid of the Chebyshev inequality). In the case of a stochastic process these theoretical values are of crucial importance as well, but since they depend on the parameter values, forecasting based on them is not as simple as in the case of a single random variable.

The expected value is, in the case of a continuous parameter,

$$E(X_t) = m(t), \quad 0 \leqslant t \leqslant T,$$

while for the discrete case

$$E(X_i) = m_i, \quad i = 1, 2, \ldots .$$

Similarly, the variance is

$$D^2(X_t) = \sigma^2(t), \quad 0 \leqslant t \leqslant T,$$

and

$$D^2(X_t) = \sigma_i^2, \quad i = 1, 2, \ldots$$

respectively.

Having a stationary stochastic process, both functions are constant:

$$E(X_t) = m, \; D^2(X_t) = \sigma^2, \; 0 \leqslant t \leqslant T,$$

$$E(X_i) = m, \; D^2(X_i) = \sigma^2, \; i = 1, 2, \ldots$$

Even knowing these quantities, drawing conclusions about future observations is not always easy because the random variables are not independent but correlated.

Here the so-called correlation function of the stochastic process plays a basic role in the theory and its applications; it is defined as follows:

$$R(t, t+s) = \frac{E[(X_t - E(X_t))(X_{t+s} - E(X_{t+s}))]}{D(X_t)\,D(X_{t+s})}\,,$$

$$0 \leqslant t < t+s \leqslant T.$$

If the process is stationary then this function does not depend on t; hence for the continuous case we obtain

$$R(s) = \frac{E[(X_t - m)(X_{t+s} - m)]}{\sigma^2}\,, \quad 0 \leqslant t < t+s \leqslant T\,,$$

and for the discrete case

$$R_k = \frac{E[(X_i - m)(X_{i+k} - m)]}{\sigma^2}\,, \quad \begin{array}{l} i = 1, 2, \ldots \\ k = 1, 2, \ldots \end{array}$$

Consider the stochastic process (33) defined in the preceding section The production process may be expected to yield a steel plate with a thickness of unchanging distribution; therefore, we hope to find a stationary process (the expected value and the variance being constant for all values of s), which together with a certain behavior of the correlation function, will ensure homogeneity (i.e., the good quality of the plate). Already having a model for the stochastic variable based on practical experiences, our aim is to estimate the theoretical quantities and to test certain hypotheses. In the case of stochastic processes the main feature of the statistical problem is that a series of observations represents only one observation of the process. Consequently, these statistical problems are more difficult to handle than in the case of a single random variable; many problems are still unsolved.

We note here that the quantity corresponding to the observed value x of a random variable X is, in the case of a random process with discrete parameter, a sequence of numbers $(x_1, x_2, \ldots)$; while for processes with continuous parameter, it is a function, the so-called *sample function*, x_t, $0 \leqslant t \leqslant T$.

As we cannot go into the details of this field here, we refer, for a general theory, to the book by Takács [A 44], by Parzen [A 32] and, concerning the statistical problems, to the book by Grenander and Rosenblatt [A 20]. In the following sections we will turn to a few special processes which often occur in industrial applications or in the practice of statistical quality control.

4.6.3 The Binomial Sequence. A Gaussian Stochastic Process

In order to illustrate the notions mentioned before we will now turn to the simplest model, the so-called binomial sequence. A continuous process, closely connected with this sequence but of a slightly different nature, will be called the binomial process.

The binomial sequence. Let the sequence

$$X_1, X_2, \ldots, X_n, \ldots$$

contain independent random variables taking on the possible values 0 and 1 with probabilities ($0 \leqslant p \leqslant 1$)

$$P(X_i = 0) = 1 - p,$$

$$P(X_i = 1) = p, \quad i = 1, 2, \ldots$$

This is one of the simplest processes, but, as the following example will show, even this is of practical interest. Having a production process, manufacturing simple products (e.g. screws or wheels) successively, let us assume that each item will be defective with probability p, independently of all previous products. The question of whether or not this assumption of independence is justified must be resolved separately and we will return to it soon.

First of all, we determine the basic quantities of the stochastic sequence. Because of this independence the joint distribution of n arbitrary elements is

$$P(X_{i_1} = \varepsilon_1, X_{i_2} = \varepsilon_2, \ldots X_{i_n} = \varepsilon_n) = \prod_{j=1}^{n} P(X_{i_j} = \varepsilon_j) =$$
$$= \prod_{j=1}^{n} p^{\varepsilon_j} (1 - p)^{1-\varepsilon_j} = p^k (1 - p)^{n-k}$$

where $\varepsilon_j = +1$ or 0 and $k = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$. As a further consequence of the independence

$$P(X_n = \varepsilon \mid X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots X_{n-1} = \varepsilon_{n-1}) = P(X_n = \varepsilon).$$

The process is obviously stationary, since in this example the independent elements are also identically distributed. For the expected value and variance we have

$$E(X_i) = E(X_i^2) = (1 - p) \cdot 0 + p \cdot 1 = p,$$

$$D^2(X_i) = E(X_i^2) - [E(X_i)]^2 = p(1 - p),$$

which holds for all $i = 1, 2, \ldots$

The correlation function is identically zero; for $i = 1, 2, \ldots$ and $k \geqslant 1$,

$$R(i, i+k) = \frac{1}{p(1-p)} E\big[(X_i - p)(X_{i+k} - p)\big] =$$

$$= \frac{1}{p(1-p)} E(X_i - p) E(X_{i+k} - p) = 0.$$

Suppose that we would like to convince ourselves of the independence of the process (e.g., that the successive piecesgo turn out to be od or defective *independently* of one another).

Assuming that we have a binomial, i.e. an independent process, let us consider the random variable

$$Y_{n,k} = X_1 X_{1+k} + X_2 X_{2+k} + \ldots + X_n X_{n+k},$$

which has the following expected value and variance:

$$E(Y_{n,k}) = np^2,$$

$$D^2(Y_{n,k}) = \begin{cases} np^2\left[1 + 2\left(1 - \dfrac{k}{n}\right)p - 3\left(1 - \dfrac{2k}{3n}\right)p^2\right], & \text{if } 1 \leqslant k < n \\ np^2(1 - p^2) & \text{, if } k \geqslant n. \end{cases}$$

If n is large enough then the standardized form

$$\omega_{n,k} = \frac{Y_{n,k} - np^2}{D(Y_{n,k})}$$

is distributed according to $N(0, 1)$ (see [B 19], p. 241).

Now we wish to test the hypothesis of independence

$$H_0: P(X_{i_1} = \varepsilon_1, X_{i_2} = \varepsilon_2, \ldots X_{i_n} = \varepsilon_N) = \prod_{j=1}^{N} P(X_{i_j} = \varepsilon_j)$$

against H_1: the X_i's are not independent. The critical region at level $1 - \alpha$ can be chosen as:

$$\mathfrak{X}_c = \{\omega_{n,k} \leqslant -u_\alpha \text{ or } \omega_{n,k} > u_\alpha\}$$

with

$$\Phi(u_\alpha) = 1 - \frac{\alpha}{2}.$$

The asymptotic normality of $\omega_{n,k}$ holds for large n and for any k. We wish to note that this test is not very efficient against all possible alternatives: a certain degree of dependence may be present and yet the random variable may still remain distributed according to the normal law with expectation 0 and with variance close to $D^2(\omega_{n,k})$. Despite this fact the test tends to indicate independence, if, after examining different k's, it turns out that $\omega_{n,k}$ falls into the acceptance region in each case.

Finally, we remark that the sum

$$s_n = X_{i_1} + X_{i_2} + \dots X_{i_n}, \quad i_j \neq i_l \text{ for } j \neq l$$

is obviously distributed, for fixed n, according to the binomial law with parameter p

$$P(s_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots n.$$

The binomial process. For the sake of completeness we turn now to a continuous process related to the preceding one. A difference between them is the following: we observe the number of certain occurrences in the time interval $(0, t)$ where the total number n of occurrences is fixed (e.g., the burning out of bulbs from a finite set which started to give light at the same time).

Let $\delta_1, \delta_2, \dots, \delta_n$ be independently and identically distributed random variables with the common distribution function

$$P(\delta_i \leqslant t) = F(t), \quad i = 1, 2, \dots, n.$$

For example, δ_i is the life span of a bulb, $X_n(t)$ is the number of δ_i's not exceeding t:

$$X_n(t) = \sum_{\delta_i \leqslant t} 1.$$

Considering $\delta_1, \delta_2, \dots, \delta_n$ as a sample with empirical distribution function $F_n(t)$, the following relation obviously holds:

$$X_n(t) = n \ F_n(t),$$

which is distributed for fixed t according to the binomial law

$$P(X_n(t) = k) = \binom{n}{k} F(t)^k (1 - F(t))^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

For the expectation and variance we have:

$$E(X_n(t)) = nF(t), \ -\infty < t < \infty,$$

$$D^2(X_n(t)) = nF(t)\,(1 - F(t)), \ -\infty < t < \infty$$

A simple calculation yields the correlation function

$$R(t, t+s) = \sqrt{\frac{F(t)}{1-F(t)}\,\frac{1-F(t+s)}{F(t+s)}}\,.$$

This means that this process is not stationary, the correlation function depending not only on s, but also on t.

If the δ_i's are uniformly distributed in the interval (0, 1), i.e.

$$F(t) = \begin{cases} 0, & t < 0, \\ t, & 0 \leqslant t \leqslant 1, \\ 1, & 1 \leqslant t, \end{cases}$$

then the process $Y_n(t) = \sqrt{n}(F_n(t) - F(t))$ has the following characteristics:

$$\left.\begin{aligned} E(Y_n(t)) &= 0, & 0 \leqslant t \leqslant 1, \\ D^2(Y_n(t)) &= t(1-t), & 0 \leqslant t \leqslant 1, \\ P(Y_n(0) = 0) &= P(Y_n(1) = 0) = 1, & \\ E[Y_n(t)Y_n(t')] &= t(1-t'), & 0 \leqslant t < t' \leqslant 1. \end{aligned}\right\} \tag{35}$$

The sample function of this process has the following shape. It starts at $t = 0$ with value $Y_n(0) = 0$ and is linear (with slope $-\sqrt{n}$) till the value $t = \delta_1^*$, which is the first-order statistic in the sample $\delta_1, \delta_2, \ldots, \delta_n$. Here $Y_n(t)$ has a jump of $\frac{1}{\sqrt{n}}$:

$$Y_n(\delta_1^* - 0) = \sqrt{n}(F_n(\delta_1^* - 0) - \delta_1^*) = -\sqrt{n}\cdot\delta_1^*,$$

while

$$Y_n(\delta_1^*) = \sqrt{n}\,(F_n(\delta_2^*) - \delta_1^*) = \sqrt{n}\left(\frac{1}{n} - \delta_1^*\right) = \frac{1}{\sqrt{n}} - \sqrt{n}\,\delta_1^*.$$

Continuing, it is again piecewise linear with the same slope $-\sqrt{n}$ and with jumps at points δ_i^*, always of the same value $\frac{1}{\sqrt{n}}$. The sample function ends at point $t = 1$, with the value $Y_n(1) = 0$ (with probability 1).

The larger n, the smaller the size $\frac{1}{\sqrt{n}}$ of the jump. If n tends to infinity, this jump tends to 0 and we obtain a limiting process $Y(t)$ with continuous sample function (with probability 1) starting at $t = 0$ with $Y(0) = 0$ and ending at $t = 1$ with $Y(1) = 0$. Since the relations (35) hold for each value of n they hold for the limiting case as well. Consequently we have for the limiting process:

$$\begin{aligned} E(Y(t)) &= 0, & 0 \leqslant t \leqslant 1, \\ D^2(Y(t)) &= t(1-t), & 0 \leqslant t \leqslant 1, \\ P(Y(0) = 0) &= P(Y(1) = 0) = 1, & \\ E(Y(t)Y(t')) &= t(1-t'), & 0 \leqslant t < t' \leqslant 1. \end{aligned}$$

For each value of $t (0 \leqslant t \leqslant 1)$ the random variable $Y(t)$ is distributed according to the Gaussian law $N(0, \sqrt{t(1-t)})$, while the joint distribution of the variables $Y(t_1), Y(t_2), \ldots Y(t_k)$, $(0 < t_1 < t_2 < \ldots < t_k < 1)$ has a k-dimensional Gaussian distribution. The continuous process described here is of the type called *Gaussian stochastic processes.*

Finally, we refer to the Smirnov and Kolmogorov distributions given in Section 3.3.3. The distribution law of Smirnov gives the probability that the sample function of the Gaussian process $Y(t)$ runs under the height y throughout the interval $0 \leqslant t \leqslant 1$:

$$P(Y(t) \leqslant y) = 1 - e^{-2y^2}, \qquad y \geqslant 0,$$

while the distribution of Kolmogorov gives the probability of the event $\{|Y(t)| \leqslant y,\ 0 \leqslant t \leqslant 1\}$.

4.6.4 The Poisson Process

Observing a certain phenomenon from time $t = 0$, a certain event occurs repeatedly at different instants. For this process the following properties will be assumed: (a) the different events occur independently of each other, (b) to a good approximation the probability of an event occurring in a very short time interval is proportional to the length of this time interval: $\lambda \Delta t$, while (c) the probability of the occurrence of more than one event during the short time interval Δt is very small compared with the probability that exactly one event occurs. Let us denote by X_t the number of events which occur during the time $(0, t)$ the distribution of which agrees with that of the number of events occurring during a time interval $(\tau, \tau + t)$. From the definition of the Poisson process, for the number of occurrences during the time t, the following distribution law can be obtained:

$$P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \ldots$$

It can be easily seen that this is a Markovian process; furthermore, it has independent increments, which means that if $t_1 < t_2 \leqslant t_3 < t_4$ then the random variables $X_{t_2} - X_{t_1}$ and $X_{t_4} - X_{t_3}$ are independent random variables. Both statements follow from the definition.

The expected value and the second moment of X_t are

$$E(X_t) = \lambda t,\ E(X_t^2) = \lambda t + (\lambda t)^2;$$

the variance is

$$D^2(X_t) = \lambda t.$$

For the correlation function of the process one obtains

$$R(\tau, \tau + t) = \frac{E[(X_\tau - \lambda\tau)(X_{\tau+t} - \lambda(\tau + t))]}{D(X_\tau)\,D(X_{\tau+t})} = \frac{1}{\sqrt{1 + \frac{t}{\tau}}}$$

i.e., this function is independent of λ. This formula can be obtained directly from the fact that the process has independent increments:

$$\begin{aligned} E(X_\tau X_{\tau+t}) &= E[X_\tau(X_{\tau+t} - X_\tau + X_\tau)] \\ &= E[X_\tau(X_{\tau+t} - X_\tau)] + E(X_\tau^2) \\ &= E(X_\tau)\,E(X_{\tau+t} - X_\tau) + E(X_\tau^2) \\ &= \lambda t\,\lambda\tau + \lambda\,\tau + (\lambda\,\tau)^2 \end{aligned}$$

since X_τ and $X_{\tau+t} - X_\tau$ are independent; now

$$\begin{aligned} R(\tau, \tau + t) &= \frac{E(X_\tau\,X_{\tau+t}) - E(X_\tau)\,E(X_{\tau+t})}{D(X_\tau)\,D(X_{\tau+t})} \\ &= \frac{\lambda\tau + \lambda t\,\lambda\tau + (\lambda\tau)^2 - \lambda\tau\lambda(t + \tau)}{\sqrt{\lambda\tau\lambda(t + \tau)}} = \sqrt{\frac{1}{1 + \frac{t}{\tau}}}. \end{aligned}$$

For fixed τ the value of the correlation function tends to zero as t tends to infinity.

Some examples of Poisson processes: the disintegration of atoms in a radioactive substance, the electron emission from the cathode of an electronic tube, telephone calls arriving at an exchange, the breakage of a certain component of a machine in a certain part of the day, the yarn breakage in a textile machine, the arrival of particles at an electron multiplier, and under certain conditions the arrival of customers to a server.

An important random variable belonging to a Poisson process is the distance between two consecutive events. Let us assume that the process (observation) starts at $t = 0$ and events occur at $0 < \tau_1 < \tau_2 < \ldots < \tau_n < \ldots$ Then τ_1, $\tau_2 - \tau_1$, $\tau_3 - \tau_2, \ldots$ are independent random variables distributed according to the exponential law with parameter λ. Let $\tau_0 = 0$; then for $k = 1, 2, \ldots$

$$P(\tau_n - \tau_{n-1} < x) = 1 - e^{-\lambda x}, \quad x \geqslant 0 .$$

From this it follows that the distribution of the random variable τ_n can be obtained in the following way. Since

$$\tau_n = (\tau_n - \tau_{n-1}) + (\tau_{n-1} - \tau_{n-2}) + \ldots + (\tau_2 - \tau_1) + (\tau_1 - \tau_2)$$

i.e., τ_n is the sum of n independent, exponentially-distributed random variables, we have (Section 2.9.5)

$$P(\tau_n < x) = \int_0^x \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}\, dt, \quad x \geqslant 0 .$$

Instead of considering this further we turn to an immediate generalization of the Poisson process which often occurs in applications.

4.6.5 The Recurrent Stochastic Process

Let us assume that events occur randomly at the consecutive time points $0 < \tau_1 < \tau_2 < \ldots < \tau_n < \ldots$ and that the differences are independent, identically distributed random variables $\tau_{i+1} - \tau_i$ $(i = 1, 2, \ldots)$ with the common distribution function $F(x)$:

$$P(\tau_{i+1} - \tau_i \leqslant x) = F(x), \quad x \geqslant 0, \quad i = 1, 2, \ldots$$

Denoting the number of events occurring during the time interval $(0, t)$ by X_t, we say that the random process X_t is a *recurrent process*. A recurrent process is, in general, non-Markovian.

For $F(x) = 1 - e^{-\lambda x}$, $x \geqslant 0$, this gives the Poisson process considered in the preceding section.

We may assume that τ_1 is distributed according to some $F_1(x)$ different from $F(x)$. We speak about a *stationary recurrent process* if τ_1 has the following distribution function:

$$F^*(x) = \frac{1}{\mu} \int_0^x (1 - F(t))\, dt, \quad x \geqslant 0$$

where

$$\mu = E(\tau_{i+1} - \tau_i) = \int_0^\infty x\,d\,F(x)\,,$$

if it is finite. In the case of the Poisson process $F^*(x)$ and $F(x)$ coincide.

In general, the distribution of τ_n cannot be obtained in an explicit form. Since τ_n is the sum of n independent random variables, the distribution of τ_n is the convolution of distribution functions, carried out repeatedly $(n - 1)$ times: Let

$$P(\tau_1 \leqslant x) = F_1(x) = \int_0^x f_1(t)dt,$$

then

$$F_2(x) = P(\tau_2 = (\tau_2 - \tau_1) + \tau_1 \leqslant x) = \int_0^x f_1(x - t)\,f(t)\,dt = \int_0^x f_2(t)dt,$$

and further

$$F_3(x) = P(\tau_3 = (\tau_3 - \tau_2) + \tau_2 \leqslant x) = \int_0^x f_2(x - t)\,f(t)\,dt = \int_0^x f_3(t)\,dt,$$

etc. Denoting the nth convolution by $F_n(x)$, i.e.

$$F_n(x) = P(\tau_n \leqslant x), \quad x \geqslant 0,$$

we are in a position to determine the distribution of X_t assuming that $F_n(x)$ is known.

The event $\{X_t \leqslant n\}$ means that the $(n + 1)$th occurrence happened after the instant t: $\{t < \tau_{n+1}\}$; conversely, the latter event is equivalent to the statement that, up to the time t, the number of events that occurred is at most n. Consequently,

$$P(X_t \leqslant n) = P(t > \tau_{n+1}) = 1 - P(\tau_{n+1} \leqslant t) = 1 - F_{n+1}(t).$$

For the expectation of X_t: $m(t) = E(X_t)$ we obtain, according to definition,

$$\begin{aligned} m(t) &= \sum_{n=0}^{\infty} n\,P(X_t = n) = \sum_{n=1}^{\infty} P(X_t = n) + \\ &+ \sum_{n=2}^{\infty} P(X_t = n) + \sum_{n=3}^{\infty} P(X_t = n) + \ldots = P(X_t > 0) + \\ &+ P(X_t > 1) + P(X_t > 2) + \ldots = \sum_{n=0}^{\infty} P(X_t > n) \\ &= \sum_{n=0}^{\infty} (1 - P(X_t \leqslant n)) = \sum_{n=1}^{\infty} F_n(t)\,. \end{aligned}$$

Let us turn to the special case when $F(x) = 1 - e^{-\lambda x}$, $x \geqslant 0$, i.e., to the Poisson process. Choosing $F_1(x) = F(x)$, we come to the stationary recurrent process, since

$$\mu = \int_0^\infty x dF(x) = \frac{1}{\lambda}$$

and

$$F^*(x) = \frac{1}{\mu} \int_0^x (1 - F(t)) \, dt = 1 - e^{-\lambda x}, \quad x \geqslant 0.$$

Furthermore

$$F_n(x) = P(\tau_n \leqslant x) = \int_0^x \lambda \frac{(\lambda \tau)^{n-1}}{(n-1)!} e^{-\lambda \tau} \, d\tau, \qquad x \geqslant 0.$$

For the expected number of occurrences during the interval $(0, t)$ we have

$$m(t) = \sum_{n=1}^\infty F_n(t) = \sum_{n=1}^\infty \int_0^t \lambda \frac{(\lambda\tau)^{n-1}}{(n-1)!} e^{-\lambda\tau} \, d\tau = \int_0^t \lambda \, (e^{\lambda\tau} e^{-\lambda\tau}) \, d\tau = \lambda t.$$

4.6.6 The Pólya Process

In this section we shall consider the discrete version of the Pólya process, i.e., the Pólya sequence. This is a stochastic sequence

$$X_1, X_2, \ldots X_n, \ldots$$

with the following properties: let $0 < p < 1$, $q > 0$, then for $n = 1, 2, \ldots$

$$P(X_i = 0) = 1 - p,$$

$$P(X_i = 1) = p \tag{36}$$

$$P(X_{n+1} = 1 \mid X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots X_n = \varepsilon_n)$$

$$= P\Big(X_{n+1} = 1 \mid X_1 + \ldots + X_n = \sum_{i=1}^n \varepsilon_i = s\Big) = \frac{p + sq}{1 + nq}. \tag{37}$$

From this we see that the case $q = 0$ leads to the binomial sequence.

Before considering some special properties of this sequence, we turn to the following example which offers an illustrative interpretation of the process. Items are produced successively on a machine. If we assume that each piece will be defective with probability p, and good with probability $1 - p$ independently, then we obtain a binomial process as considered in Section 4.6.3. But let us suppose now that we no longer demand independence and turn to a model which is closer to reality. In many cases the following assumption presents itself: if a piece turns out to be good then we have a larger probability for the next piece to be good and conversely it is more probable a defective item will be followed by another defective item rather than by a good one. If this qualitative assumption is formulated quantitatively according to (37), we come to the Pólya sequence.

As can be seen, the covariance function is

$$E[X(_i - p)(X_{i+k} - p)] = p(1-p)\frac{q}{1+q}$$

independent of i and k. Together with (36) this suggests that the Pólya sequence is a stationary stochastic process, which can be shown to be the case.

For the distribution of the sum $\varkappa_n = \sum_{i=1}^{n} X_i$, i.e., for the number of defectives in n (consecutive) pieces, the following holds:

$$P(\varkappa_n = k) = \binom{n}{k} \frac{\prod_{i=0}^{k-1}(p+qi)\prod_{j=0}^{n-k-1}(1-p+jq)}{\prod_{l=0}^{n-1}(1+lq)}, \quad k=0, 1, 2, \ldots n.$$

The same is true for any sum $\sum_{j=1}^{n} X_{i_j}$, irrespectively of the choice of the indices if the i_j's are different. This means that it is true for a randomly chosen sample of size n taken from a lot. Again, taking $q = 0$ we arrive at the binomial law:

$$P(\varkappa_n = k / q = 0) = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, 1, 2, \ldots, n.$$

Let us suppose now that in a lot containing N items the number of defective items is M. Then, in a sample of size n ($n < M$) taken randomly from this lot, the number $\varkappa_n$ of defective pieces follows the following (hypergeometric) distribution:

$$P(\varkappa_n = k \mid \varkappa_N = M) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \ldots n.$$

This expression can be rewritten in the following form:

$$\binom{n}{k}\frac{M(M-1)\ldots(M-k+1)(N-M)(N-M-1)\ldots(N-M-n+k+1)}{N(N-1)(N-2)\ldots(N-n+1)}$$

$$=\binom{n}{k}\frac{\prod_{i=0}^{k-1}\left(\frac{M}{N}-i\frac{1}{N}\right)\prod_{j=0}^{n-k-1}\left(1-\frac{M}{N}-j\frac{1}{N}\right)}{\prod_{l=0}^{n-1}\left(1-l\frac{1}{N}\right)}$$

which shows that the Pólya sequence leads to the hypergeometric distribution with $p=\frac{M}{N}$ and $q=-\frac{1}{N}\cdot$

CHAPTER III

METHODS OF STATISTICAL QUALITY CONTROL

1 STATISTICAL METHODS IN THE CONTROL OF PRODUCTION PROCESSES

1.1 GENERAL PART

1.1.1 Quality Control and Statistical Quality Control

The use of mathematical methods in quality control has the following advantages:

1. It gives the possibility of a *more efficient*, and therefore, more economical, process control and final acceptance inspection.
2. It helps improve the quality and the uniformity of the products.
3. It provides for the supervision and development of technological procedures and equipment.

In non-statistical control, however, conclusions based on a sample are not scientifically justified. Often both the choice of the sample size and the inference drawn from the sample are left to the controller. There is a considerable danger of that this person is predisposed to identify the sample with the entire sampled population. If, for example, there is no defect in one sample, he is inclined to regard the whole population as defectless.

In contrast, statistical quality control makes it possible to utilize statistical theory to draw the most accurate conclusion possible from the sample. The statistical methods assure a relative safety on the basis of a limited amount of control; this means a considerable economic advantage.

As mentioned in I. 1.2, the two main forms of statistical quality control are *control during production* (or production process control) and *acceptance sampling* (product control). The subject matter of the present section is the former; section 2 will deal with the latter.

Let us note that this classification is made in reference to the fields of application.

The respective methods developed in connection with these applications are not necessarily restricted. They—as well as other statistical methods (regression analysis, etc.)—have important roles in various places where quality control works, and in the up-to-date total quality assurance programs of the industrial plants.

Apart from the general advantages of statistical methods already mentioned, statistical process control has further particular advantages over the nonstatistical methods. In many cases it is able to detect the deterioration of the production process even before any defective item appears in the sample. Often it can detect trends or other changes occurring in the process, and by this, influence the process and prevent the production of defects.

Usually, process control is performed by means of *control charts.* The observed characteristics of samples, taken at regular intervals in the course of production, are plotted on these charts. The control charts are constructed in such a way that the result of the statistical test to be performed on the sample (II. 4.3.1) can be seen directly. In addition, they give a dynamic survey of the whole process. The control chart can be regarded as the documentation of the sampled process and can be utilized at any later time if the supervision of the manufacturing process is necessary.

The proper choice of the null hypothesis, tested repeatedly with the aid of the control chart, is of great importance. Not only are the given specifications of the products to be taken into account. The economic use of the sample method—at customary sample sizes and sampling invervals—usually requires that the null hypothesis should have a reasonably large a priori probability; this means that the null hypothesis must reflect the normal state ("state of control") of the production process. The best way of examining the compatibility of these two conditions is by means of a *preliminary analysis* based on a *preliminary collection of data.* If the process turns out not to meet the specifications, then either the specifications of the manufacturing process (machine, raw material, etc.) should be revised or some other appropriate action must be taken. If, however, the process is satisfactory, then this preliminary sampling will serve as the basis for selecting an appropriate null hypothesis.

Usually, the preliminary analysis is considered a part of the statistical process control. However, they are often used separately: sometimes only the data analysis is performed; in other cases the control charts are introduced without any preliminary sampling.

The control chart method was introduced by W. A. Shewhart in 1924 [B 71].

1.1.2 Different Types of Process Control

As mentioned in I. 1.2, the two main types of process control are the following: control by variables and control by attributes.

If the items are investigated for qualitative characteristics, only control by attributes can be used. If, however, some quantities are measured and recorded, there is the possibility of using quantitative statistical methods, i.e., control by variables.

In some cases qualitative statistical methods are applied even if the characteristics in question are quantitative. This may occur, e.g., under the following circumstances:

(a) Lack of appropriate measuring instruments. If the quantitative characteristic is controlled by gages, only qualitative statistical methods can be applied. Moreover, if the precision of the measuring instrument is not sufficient (the smallest distinguishable unit is more than one tenth of the range of variation or of the tolerance), it is impossible to use a quantitative method.

(b) Even if the quantitative characteristics are measured with sufficient precision, we may decide to use control by attributes, for reasons of economy. This may be the case when many quantitative characteristics are measured on each item, and it is not possible to select the one or two most important ones.

From the statistical point of view, a method is classified as being *qualitative* (*control by attribute*) if it is based on no more than the number of sample units which fail to conform to some requirements, or on the number of defects found in the sample (one unit may have more defects). The method is *quantitative* (*control by variables*) if some arithmetic operations are performed on the quantitative data (e.g., computing averages) and their results are used. The main advantage of the method of variables lies in the more complete use of the information contained in the sample. In the majority of cases where both methods are possible, this makes it worthwhile to perform the additional work required.

In process control, the particular advantages of control by variables are the following:

(a) The simultaneous control of the mean level and the variation of the controlled variable is possible.

(b) The difference in labor required by the two methods is smaller than in acceptance sampling.

The most important control charts for variables are the chart for median and range, and the chart for mean and range, respectively.

Among the control charts for attributes, the chart for the number of defectives is the most important.

1.1.3 The Outlines of Statistical Process Control

As already mentioned in 1.1.1, the control of the production process consists of two major steps: a *preliminary analysis* (resulting in some decisions, as will be seen below); and the process control itself in the narrow sense. The best way to assure all the benefits of the control chart method is to perform the steps successively.

The purposes of preliminary analysis are the following:

(a) To obtain information on the unknown process, e.g., on the distribution of the quality characteristic to be examined, on whether the machine to be used is appropriate for the planned purpose. To settle the adjustment of the machine, the characteristics to be checked, etc.

(b) To serve as a basis for the control chart to be used in the course of the production process.

In some cases, however, the analysis may serve the purposes listed in (a) only, and will not be followed by the use of the control chart method. In this case we speak simply about *process analysis*.

The control chart method means that the process is controlled by samples taken repeatedly at certain time intervals, and conclusions drawn whether the process is still under control or not. If necessary, appropriate actions will be taken for assuring the maintenance of good quality.

1.2 PROCESS CONTROL BY VARIABLES

1.2.1 Process Analysis

Before starting a production process one has to know what the machine or manufacturing equipment can do. In the majority of cases the most important question concerns the precision of production. In order to be able to give a reliable answer to this question we have to perform a process analysis.

Let us now point out that, apart from the machine, many circumstances may influence the production process. If a new process is started the effects of which are not yet known, it is advisable to perform a process analysis at the beginning. The results of this analysis help to make a final decision on whether to leave the process unchanged, or to alter the machine or the circumstances, or to take some other appropriate action. If the process goes on unchanged, the control chart, designed on the basis of the preliminary analysis, will help us to hold the quality constant. The data obtained in the process analysis could be evaluated by the method described in II. 3.2. But this approach does not take into account the chronological order of the observations and therefore cannot detect a shift or other variation in time. This is a disadvantage of this form of analysis; such variations should almost always be considered and it is desirable that the analysis detect them.

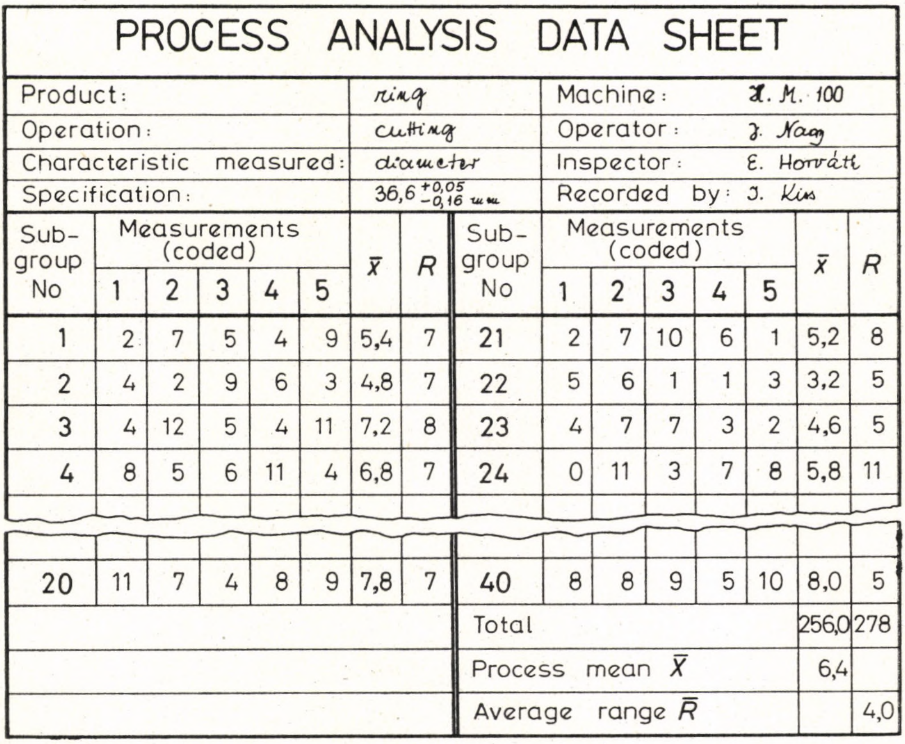
Therefore, in general, the data are divided into subgroups according to the chronological order of production, and the within-subgroup variation is estimated. If the underlying distribution is normal, the standard deviation may be estimated by the *mean range* (average of the subgroup ranges). This method of evaluation is dealt with in the forthcoming sections.

1.2.1.1 Techniques for the Collection of Data. For the purposes of process analysis it is desirable to take a sample of at least 200–300 items. If we aim at investigating a newly started process then, in general, it is advantageous to select the first 200–300 manufactured items. This diminishes the risk that some irregular event will disturb the analysis.

It is advisable to use special forms for recording, representing and evaluating the data. Such forms are shown in Fig. 36 (Data Sheet) and Fig. 37 (Data Diagram). The data have been listed in chronological order in subgroups of five on the data sheet, from which they have been transferred onto the data diagram.

The formation of subgroups is advantageous from the point of view of graphical representation and evaluation of the data, even if the whole sample consists of items manufactured consecutively. If our analysis precedes the introduction of a control chart method, then the group size should, if possible, agree with the sample size we intend to use for the control chart. The most frequently used subgroup size is five.

For convenience, the measurements are usually coded; this is the situation in our Figures as well, where the differences from the nominal value are recorded and the unit is chosen as the smallest unit to be read off on the measuring instrument. The origin of the code scale may be the nominal value (as it is in our example) or any other suitably chosen value.



PROCESS ANALYSIS DATA SHEET

Product:	ring	Machine:	X. M. 100
Operation:	cutting	Operator:	J. Nagy
Characteristic measured:	diameter	Inspector:	E. Horváth
Specification:	$36{,}6^{+0{,}05}_{-0{,}16}$ mm	Recorded by:	J. Kiss

Sub-group No	Measurements (coded)					$\bar{x}$	R	Sub-group No	Measurements (coded)					$\bar{x}$	R
	1	2	3	4	5				1	2	3	4	5		
1	2	7	5	4	9	5,4	7	21	2	7	10	6	1	5,2	8
2	4	2	9	6	3	4,8	7	22	5	6	1	1	3	3,2	5
3	4	12	5	4	11	7,2	8	23	4	7	7	3	2	4,6	5
4	8	5	6	11	4	6,8	7	24	0	11	3	7	8	5,8	11
20	11	7	4	8	9	7,8	7	40	8	8	9	5	10	8,0	5
								Total						256,0	278
								Process mean $\bar{\bar{x}}$						6,4	
								Average range $\bar{R}$							4,0

FIG. 36.

While the data are collected, homogeneity of the production process must be assured: the possibility of disturbing effects has to be eliminated as far as possible. The circumstances of the production process and those of the control of the items are to be specified in advance. Any change which may influence the process or the measurements is to be recorded on the data sheet.

The measuring instrument should enable us to read off at least 10 different values between the specified limits.

For the sake of simplicity, in certain cases the use of the data sheet may be omitted and the data introduced immediately onto the data diagram. In this case the head of the data sheet, as shown in Fig. 36, is to be transferred onto the top of the data diagram.

In this Diagram (Fig. 37) the spaces between the consecutive horizontal lines, rather than the lines themselves, correspond to the marked numerical values. This is suitable if the number of different possible values

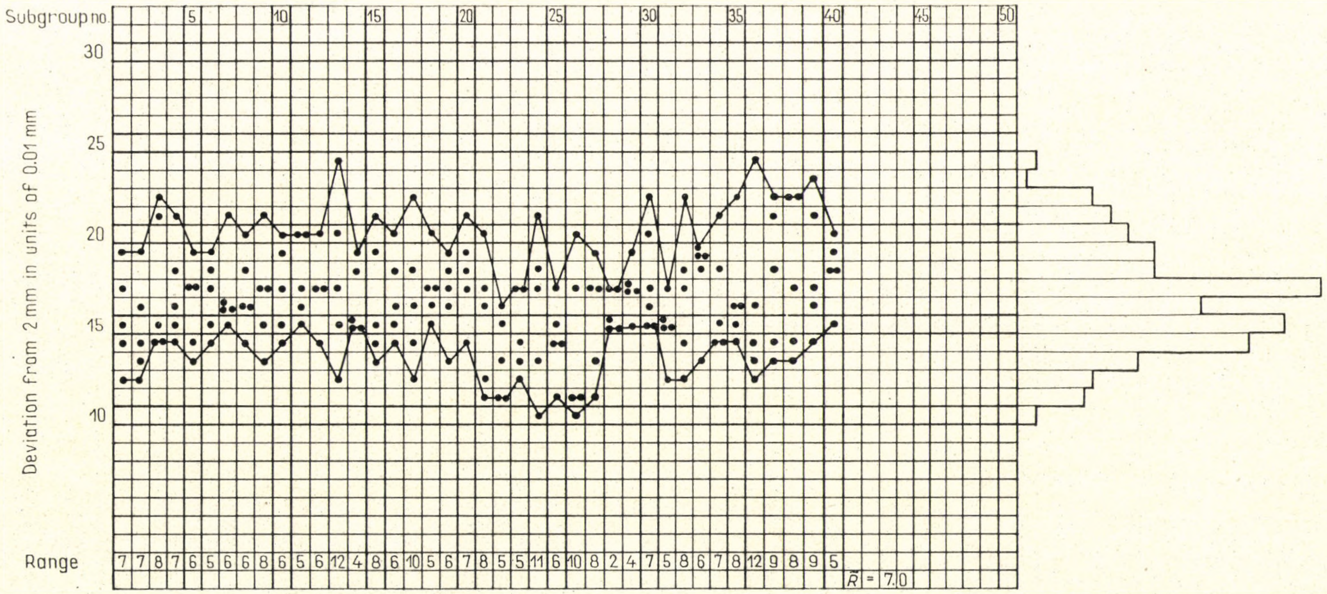


FIG. 37. Process analysis data diagram

is not too large. Accordingly, the points representing items are plotted between the lines; if equal values occur within the same group, several points fall into the same square (as for example is the case, in Fig. 37, in the subgroup No. 7).

1.2.1.2 Evaluation of the Data and Calculation of Statistical Characteristics. As can be seen in Fig. 36, the form facilitates the computations of the average $\overline{\overline{X}}$ of all data (grand average) and of the mean range $\overline{R}$ (the average of the subgroup ranges). If systematic changes are fore- seen or an adjustment must be made after finishing the analysis and before starting the control chart, then the calculation of $\overline{\overline{X}}$ is not necessary. If no data sheet is used, the ranges can be read off and recorded on the data diagram as shown in Fig. 37. The mean range multiplied by the factor $1/d_2$ of Table 8 (where k is the group size) gives an estimation of the standard deviation of the process, which is unbiased if the original data have a normal distribution (d_2 is the expectation of the range of a sample of size n taken from the standardized normal distribution). This remains true even in the case of systematic changes in the mean value, i.e., if the elements of all subgroups can be considered as normally

TABLE 8

ESTIMATION OF VARIANCE AND PROCESS CAPABILITY FROM $\overline{R}$
(normally distributed variables)

Subgroup size n	d_2	$\frac{1}{d_2}$	$H = \frac{6}{d_2}$
2	1.128	0.886	5.32
3	1.693	0.591	3.54
4	2.059	0.486	2.91
5	2.326	0.430	2.58
6	2.534	0.395	2.37
7	2.704	0.370	2.22
8	2.847	0.351	2.11
9	2.970	0.337	2.02
10	3.078	0.325	1.95
11	3.173	0.315	1.89
12	3.258	0.307	1.84

distributed variables with a common variance, but with different expectations in subgroups.

One of the main advantages of the formation of subgroups is the reliability of the estimation procedure even in the case of a shift or another change in the central tendency, shifts being very frequent in production processes.

In practice, in the majority of cases the estimator $\bar{R}/d_2$ is not computed, since all quantities needed can be calculated directly from $\bar{R}$ with the help of Table 8.

1.2.1.3 Determination of the Accuracy of the Machine. The process analysis aims—among other things—to determine the limits at which the machine is able to work, provided that the process is not disturbed by any assignable cause. Usually, we estimate the *capability* or "*natural tolerance*" of the process, i.e., the distance between the mentioned limits, by multiplying the estimate of the standard deviation by 6. We know that in the case of a normal distribution, an interval of this width covers practically the whole distribution, i.e., the probability that an observation lies outside a centered interval of this length is only 0.27%. Departure from normality and estimation error may increase this probability considerably, but in the majority of practical cases it is still negligible.

If the mean range $\bar{R}$ has already been calculated as shown in the last Section above, the process capability can be immediately (i.e., without estimating the standard deviation) estimated by multiplying $\bar{R}$ by the constant $H = 6/d_2$. The values of this constant are given in Table 8.

If the underlying distribution is not normal, this estimator is not unbiased. Experience, however, shows that even in these cases the resulting interval may, for all practical purposes, be considered to be the spread of the distribution.

If the collected data cannot be covered by an interval of length $H\bar{R}$ (3–4 exceeding units out of 2–300 are still acceptable) then either assignable causes of variation (trend, jump, etc.) are present or the distribution differs sharply from the normal law.

Trend (shift in time) is very frequent in production processes. Even if the data obtained in the process analysis belong to consecutive items, as recommended in III. 1.2.1.1, the time needed for collecting 2–300 items may suffice to demonstrate a shift in the process mean. If such a shift or another variation (jump, slow fluctuation) is apparent from the data diagram, efforts should be made to eliminate its cause; if this is not possible then we have to conclude that this variation must be reckoned with in the production process.

If the variation in time is not large enough to be apparent from the data diagram, the underlying distribution can be examined by representing it in a histogram. This is done in Fig. 37 (on the right); the frequencies have been simply counted in the horizontal rows. We may decide on the basis of the shape of the histogram whether the distribution may be regarded as normal. Alternatively, we may turn to a more accurate method (II. 4.3.7).

If a considerable trend or some other variation is observed in the course of the time interval covered by the process analysis, then a histogram constructed in the above way would represent a mixed distribution. The fact that this mixed distribution appears to reflect the normal law does not guarantee that the momentary distribution approximates it also, and vice versa. The shape and spread of the mixed distribution depends on the time covered by the observations. If investigation of the momentary distribution is desired, the deviations from the smoothed trend line or from the group averages can be used. For more exact tests see II. 4.3.7.3.

If the distribution is sharply different from the normal one, then the natural tolerance may be estimated from the histogram, or its calculation may be based on some other specified theoretical distribution.

Instead of, or in addition to, the above investigations, the presence of disturbing effects and their importance from the point of view of the control chart method may be tested by plotting the data in a control chart in which the control limits are calculated from the same data. If no point (or perhaps a single one) falls outside the control limits, we say the process is in control.

The next step is to compare the process capability $H\bar{R}$ with the actual specification limits. Let us first consider the case where two specification limits are given and the process mean can be set up by adjustment. (Figure 38.)

(a) If the process capability is less than the distance between the specified limits, defect-free production can be maintained if the process is centered properly. In this case we have a latitude in the adjustment of the central level. This is usually used to prevent the undesired effect of a possible shift in the process average. We will return to this question in the next Section.

(b) If the process capability is approximately equal to the distance between the upper and lower specification limits, then defectless production is possible if, and only if, the process is centered in the middle of the specification limits. Once this has been achieved, all possible steps are to be taken to maintain the centering. If possible, by reducing the process dispersion by some appropriate action, or by broadening the tolerances.

(c) If the difference between the specification limits is definitely less than the process capability, then defect-free production is not possible. Efforts must be made to decrease the process dispersion and/or to review the tolerances. Also, the possibility of producing the product on another machine is to be considered. If the original decision to use the first machine is maintained, then it is very important to center the process in the middle of the specification limits. By doing so, the proportion of defective items will be minimized.

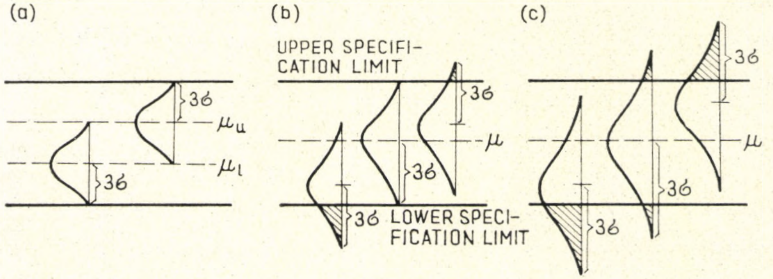


FIG. 38. Tolerance and process capability

(d) The case of a single specification limit is similar to case (a). In this case an appropriate machine setting will leave a margin of safety between the specification limit and the "natural tolerance limit".

Let us remark that the process capability is a quantity characterizing the machine together with its environment; therefore it must be recorded for possible future reference when decisions affecting the machine are to be made.

1.2.2 Adjustment of the Machine

If the data analysis does not show a considerable trend in the mean, and the values of the estimated constants assure us that the natural tolerance lies between the specification limits, then no adjustment is necessary.

In many cases, however, the desired average must be achieved by adjustment of the machine. The viewpoints for determining the best level of the process average have been indicated in the last Section above. So in the cases (b) and (c) the most suitable average value is the central line between the specification limits—provided that there is no rapid trend. In the case of (a) it is suitable to take the shift direction into account and, accordingly, to assign the process average at a certain distance from the middle of the specification limits.

The adjustment is always to be controlled. However it is not sufficient to investigate a single item produced after the adjustment. At least a sample of the size used in the control chart should be taken (in this case we can start the control chart with this sample). In some cases even larger samples may be necessary. Also the application of the Burr–Shainin Diagram for Machine Setting may be considered [B 13].

Example. Starting from the data of Fig. 36 (Data Sheet) or Fig. 37 (Data Diagram) we obtain for the value of the mean range $\bar{R} = 7$.

For the process capability we get with the help of Table 8

$$H\bar{R} = 2.58 \cdot 7 = 18.$$

Since the specification limits are 10 and 20, the process capability is broader than tolerance and defect-free production is not possible. To minimize the defect the process should be centered at the mean value $\mu = 15$.

1.2.3 Statistical Control During the Production Process

1.2.3.1 Control Charts. The basic principle of the control chart has been given in 1.1.1. It is a tool for a repeated application of a statistical test on the same hypothesis (it is desirable that this hypothesis be based on the data of the preliminary analysis).

Before starting the control chart we must decide on the significance level, the sample size, the frequency of sampling and the type of chart to be used.

In the special literature of quality control the term *probability level* is used for the level of the test (II. 4.3.1 and 4.3.2). This is the complementary probability of the significance level. For example, if the probability level is 99.8% (in this case the control limits are called 99.8% probability limits), the significance level is 0.2%.

The proper selection of the probability level is basically a question of the economical consequences of both kinds of errors. Limits that are too broad may give insufficient protection against deterioration of the process; excessively narrow limits may cause frequent set-ups, even if the process fits the specifications. Difficulties in determining an optimal level have been mentioned in II. 4.3.2.

The most commonly used probability level is 99.8% or some value near it. The "3-sigma control limits", used in the American standards, approach this value for the mean chart (in fact the probability that the sample mean lies within the 3-sigma limits is 99.73%). For other types of charts the probability level of the 3-sigma limits is lower than 99.73%, and it varies with the type of chart and the sample size. For the range chart and standard deviation chart it is approximately 99%.

Similarly, the actual probability of error of the first kind is undetermined if 3-sigma limits or control limits of fixed probability level based on the normal law are used in the case of an unknown distribution.

In this book, tables for calculating 99.8% probability limits are given.

Sometimes the use of limits considerably broader or narrower than 99.8% may be justified. Some sources, including British Standards, give the corresponding constants for 98% and 95% on control charts. The distance of the 99% probability limits from the central line is nearly five sixths that of the 99.8% limits. The corresponding rate between the 95% and 99.8% probability limits is nearly two thirds.

The most commonly used sample size is five. In individual cases, larger or smaller sample sizes may sometimes be suitable. The larger the sample, the more sensitive the control chart will be to shift. The increase of the sample, however, frequently causes practical difficulties.

In problems related to the probability level and the sample size, information yielded by the power function (or OC function) may be taken into account. The OC functions of different types of control charts have been studied, e.g., by Duncan [A 12], pp. 305–315; and Cowden [A 7], pp. 257–282.

The problem of the frequency of sampling will be dealt with in Section 1.4.

The most commonly used types of control charts for variables are the "Chart for Median and Range" (M and R Chart, Section 1.2.3.3) and the "Chart for Mean and Range" ($\overline{X}$ and R Chart, Section 1.2.3.2). The sample mean gives better protection against shift than the median, whereas the advantages of the latter are: (1) its simplicity, and (2) more visual information, since all observations are plotted into the diagram.

Under usual circumstances, the range chart gives sufficient protection against an increase of dispersion. Only in exceptional cases is the use of the more precise method, the Chart for Standard Deviation (Section 1.2.3.4), worth the extra work connected with it.

Some other types of control charts are dealt with in Section 1.2.3.4.

1.2.3.2 Chart for Meanand Range. On the chart for mean and range ($\overline{X}$ and R chart) two diagrams are plotted: one for the mean $\overline{X}$ and one for the range R. In this way, the centering and the variability of the process are simultaneously controlled (Fig. 40, p. 247).

The computation of the control limits of the mean is to be performed according to the formulas

$$UCL_{\overline{X}} = \mu + A_2\overline{R}$$

$$LCL_{\overline{X}} = \mu - A_2\overline{R}$$

where $UCL_{\overline{X}}$ and $LCL_{\overline{X}}$ denote the upper and lower control limits, respectively, and μ is the process average known from preliminary analysis, from other past data, or from adjustment at the start of the control chart [if the control chart is used for analyzing past data (cf. Section 1.2.1.2), then μ is estimated by the average of all data]; $\overline{R}$ denotes the mean range calculated from the preliminary analysis; $A_2 = A/d_2 = u/(d_2\sqrt{n})$ where n is the sample size, d_2 is the expectation of the range of a sample of size n taken from the standardized distribution, u is the upper α quantile of the standardized normal distribution if 2α is the chosen significance level. In principle, these formulas can be used if the shape of the underlying distribution is known but differs from the normal distribution. (Remember that for practical purposes the sample mean can always be considered normally distributed.) However, we usually work with a normal parent distribution; accordingly, d_2 denotes the expectation of the range of a sample taken from a standardized normal distribution, tabulated in Table 8, and u is the $1 - \alpha$ quantile of the standardized normal distribution (i.e., $\Phi(u) = 1 - \alpha$); u can be determined from Table IV (p. 385).

The values of u, and therefore those of A and A_2, depend on the prescribed probability level. If it is desirable to indicate the probability level in the notations, then $A = A_\alpha$ or $A_2 = A_{2(\alpha)}$ may be used (for example $A_{.001}$ and $A_{2(.001)}$). In Table 9 they are tabulated for the probability level 99.8%, i.e., we might write the more complete notation $A = A_{.001}$ and $A_2 = A_{2(.001)}$. (Note that some sources use the shorter symbols A and A_2 for the "three-sigma" limits only.)

For other probability limits, the corresponding constants may be calculated with the help of Table IV.

TABLE 9

FACTORS FOR CONTROL LIMITS
Probability level: 99.8%

Sample size	A_2	D_3	D_4	F	A	D_1	D_2	B_2
2	1.95	0.00	4.12	2.92	2.18	0.00	4.65	2.33
3	1.05	0.04	2.98	1.23	1.78	0.06	5.06	2.15
4	0.75	0.10	2.57	1.09	1.54	0.20	5.31	2.02
5	0.59	0.16	2.34	0.72	1.38	0.37	5.48	1.92
6	0.50	0.21	2.21	0.70	1.26	0.54	5.62	1.85
7	0.43	0.26	2.11	0.53	1.17	0.69	5.73	1.79
8	0.38	0.29	2.04	0.53	1.09	0.83	5.82	1.74
9	0.35	0.32	1.99	0.43	1.03	0.96	5.90	1.70
10	0.32	0.35	1.93	0.44	0.98	1.08	5.97	1.67
11	0.29	0.38	1.91	0.36	0.93	1.20	6.04	1.64
12	0.27	0.40	1.87	0.37	0.89	1.30	6.09	1.61

The constant A can be used for computing control limits if the standard deviation σ has been estimated (by any method). In this case the formulae

$$\begin{aligned} UCL_{\overline{X}} &= \mu + A\sigma \\ LCL_{\overline{X}} &= \mu - A\sigma \end{aligned} \tag{1}$$

are valid.

The limits for the range are to be calculated from the equations:

$$\begin{aligned} UCL_R &= D_2\overline{R} \\ LCL_R &= D_1\overline{R} \end{aligned}$$

where, again, UCL_R and LCL_R denote the control limits of the range; the constants D_1 and D_2 are to be read from Table 9. Again, our Table has been prepared according to the principle of fixed probability limits, i.e., $D_2 = w_2/d_2$ and $D_1 = w_1/d_2$, where w_2 and w_1 are the .999 and .001 quantiles of the distribution of the range of a sample from a standardized normal distribution; d_2 is the expectation of the same range. Naturally, these constants depend on the sample size. In some places they are given for different probability levels (e.g. British Standards). If it is desirable to indicate the level (quantile index) in the notation, or to indicate that the constants are not used for calculating "three-sigma" limits, we use the notations $D_{.p}$, $w_{.p}$ or, e.g. in our case, the notations $D_{.001}$, $D_{.999}$, $w_{.001}$, $w_{.999}$, instead of D_1, D_2, w_1 and w_2.

According to the principles of the three-sigma limit system, $D_2 = 1 + 3e_2/d_2$, $D_1 = \max(0, 1 - 3e_2/d_2)$, where e_2 is the standard deviation of the mentioned range distribution. Tables giving such constants are to be found, e.g., in Juran [A 27], p. A/18.

The constants $w_{.999}$ and $w_{.001}$ can be used for calculating control limits from the population standard deviation σ.

μ and $\bar{R}$ can be marked on the diagrams as central lines for the means and ranges, respectively. If they are used, they are indicated by fainter lines than the control limits; usually they serve to inform the supervisor and not the controller.

Sometimes in the R chart the lower control limit is omitted. Although a point falling below this limit may be taken as a sign for assignable change—this change being favorable—no action is required. In this case, however, the probability of an error of the first kind decreases, i.e., the actual probability level is 99.9% if UCL_R is computed by the above formula.

In the computation of the limits, it is sufficient to retain one more decimal than in the original readings.

Even if the underlying distribution is not normal, the formulae (1) can be used for calculating the control limit of the mean, provided we have an appropriate estimation of the population standard deviation σ. For practical purposes, the distribution of the mean can be considered normal in this case also. Special calculations may be performed for computing control limits for the range, or control limits calculated on the basis of the normal assumption may be modified according to experience.

Once the control limits have been calculated and plotted on the chart, the control-chart-guided control may start. The data of the samples, periodically taken, may be first introduced onto a data sheet (Fig. 39), where the sample means and ranges will be computed. These results are to be transferred onto the chart. In the chart the ordinates represent the mean and the range of the sample, respectively, and the abscissa represents time (Fig. 40). Usually, coded scales are used (cf. 1.2.1.1).

If a point falls outside the limits, this is to be regarded as an indication of an out-of-control state. The technical cause is to be investigated, and appropriate action is to be taken (see, e.g., Juran [A 27], pp. 13/57—58).

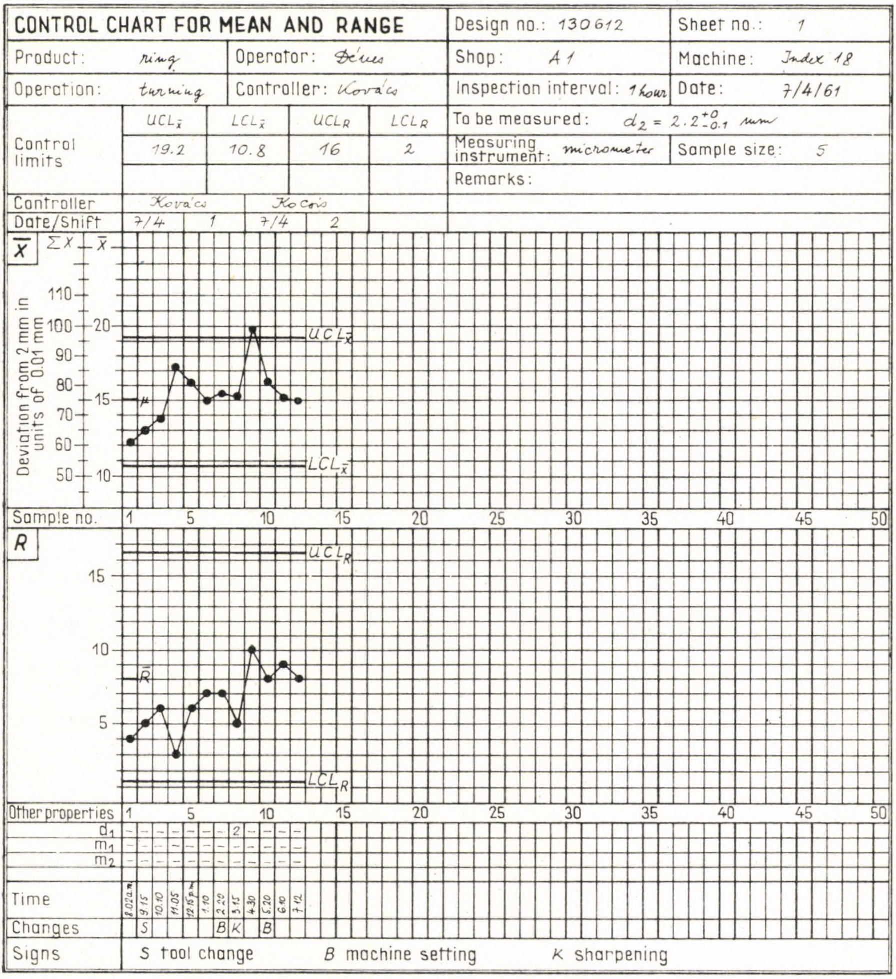
DATA SHEET
For control of mean and range

Design No: *452 A*
Product: *ring*
Specification limits: $\genfrac{}{}{0pt}{}{2.27}{2.05}$ *mm*
Code: *Deviation from 2 mm in units of 0.1 mm*

Machine: *Index 18*
Operator: *Dénes*
Controller: *Kovács*
Inspection interval: *1 hour*

Sample No	Date	Time	Measurements x					$\bar{x}$	R	Sample No	Date	Time	Measurements x	$\bar{x}$	R
1	*7/4*	*8.02 a.m.*	*12*	*12*	*11*	*15*	*11*	*61*	*4*	26					
2		*9.15*	*13*	*16*	*11*	*13*	*12*	*65*	*5*	27					
3		*10.10*	*18*	*13*	*12*	*12*	*13*	*68*	*6*	28					
4		*11.05*	*16*	*16*	*18*	*17*	*19*	*86*	*3*	29					
5		*12.15 p.m.*	*19*	*18*	*13*	*15*	*16*	*81*	*6*	30					
6		*1.10*	*18*	*12*	*14*	*19*	*12*	*75*	*7*	31					
7		*2.20*	*12*	*18*	*13*	*17*	*17*	*77*	*7*	32					
8		*3.15*	*15*	*17*	*13*	*13*	*18*	*76*	*5*	33					
9		*4.30*	*13*	*20*	*21*	*23*	*22*	*99*	*10*	34					
10		*5.20*	*16*	*21*	*17*	*14*	*13*	*81*	*8*	35					
11		*6.10*	*17*	*12*	*18*	*19*	*10*	*76*	*9*	36					
12		*7.12*	*13*	*11*	*19*	*16*	*16*	*75*	*8*	37					
13										38					
14										39					
15										40					

FIG. 39. Data sheet for $\overline{X}$ & R chart
(Filling in is exemplified by *Italics*)



CONTROL CHART FOR MEAN AND RANGE

Design no.: 130612	Sheet no.: 1		
Product: ring	Operator: Déves	Shop: A1	Machine: Index 18
Operation: turning	Controller: Kovács	Inspection interval: 1 hour	Date: 7/4/61

To be measured: $d_2 = 2.2^{+0}_{-0.1}$ mm

Measuring instrument: micrometer

Sample size: 5

Remarks:

Control limits	$UCL_{\bar{x}}$	$LCL_{\bar{x}}$	UCL_R	LCL_R
	19.2	10.8	16	2

Controller	Kovács		Kocsis	
Date/Shift	7/4	1	7/4	2

Sample no.	1	2	3	4	5	6	7	8	9	10	11	12
d_1	–	–	–	–	–	–	–	2	–	–	–	–
m_1	–	–	–	–	–	–	–	–	–	–	–	–
m_2	–	–	–	–	–	–	–	–	–	–	–	–
Time	8.02 a.m.	9.15	10.10	11.05	12.15 p.m.	1.10	2.20	3.15	4.30	5.20	6.10	7.12
Changes		S					B	K		B		

Signs: S tool change B machine setting K sharpening

$\bar{X}$
ΣX
$\bar{x}$
Deviation from 2 mm in units of 0.01 mm
110
100
90
80
70
60
50
20
15
10
μ
$UCL_{\bar{x}}$
$LCL_{\bar{x}}$
Sample no.
1
5
10
15
20
25
30
35
40
45
50
R
15
10
5
$\bar{R}$
UCL_R
LCL_R
Other properties

FIG. 40.

If the sample size varies, then either the control limits are to be recalculated each time, or—if the fluctuations are not large—we can calculate the control limits for the average sample size.

Apart from analysis based on statistical theory, the configuration of successive points may give visual evidence of a shift of the process mean. Warning limits or cumulative sum charts (Section 1.2.3.4) may be extremely useful in the hands of people who understand control chart theory. However, in the majority of cases the simpler types of control charts, such as those discussed in the present and the next paragraphs, are more suitable. Frequently, the quality control engineer, or another qualified person, takes into consideration the configuration of the points, without the use of strict theoretically based rules. He may conclude that assignable changes in the production areas are present, without any control limits having been exceeded. It is desirable, however, that the person charged with this task be familiar with the theory of the more sophisticated types of control charts.

In some cases when the process dispersion is stable, the range control diagram may be omitted and only the *control chart for the mean* must be maintained. An intermediate solution consists of recording the ranges without using a diagram form.

A special case of the control chart for the mean is the *control chart for individual measurements*. In this case each sample consists of a single element.

In some other special cases only the *control chart for the range* is maintained.

Modified control limits. If the distance between the specification limits is definitely greater than the process capability (cf. case (a) in Section 1.2.1.3) the use of the $\bar{X}$ and R chart with *modified control limits* may be suitable. In this case we allow the process average to move between two limits $\mu_l < \mu_u$. If the underlying distribution is near the normal law, we usually use

$$\mu_u = USL - 3\,\sigma \approx USL - \frac{H}{2}\bar{R}; \quad \mu_l = LSL + 3\,\sigma \approx LSL + \frac{H}{2}\bar{R}$$

where USL and LSL are the upper and lower specification limits, respectively. Now the upper control limit is to be calculated from μ_u and the lower one from μ_l, i.e.

$$UCL_{\bar{X}} = \mu_u + A_2\bar{R}$$

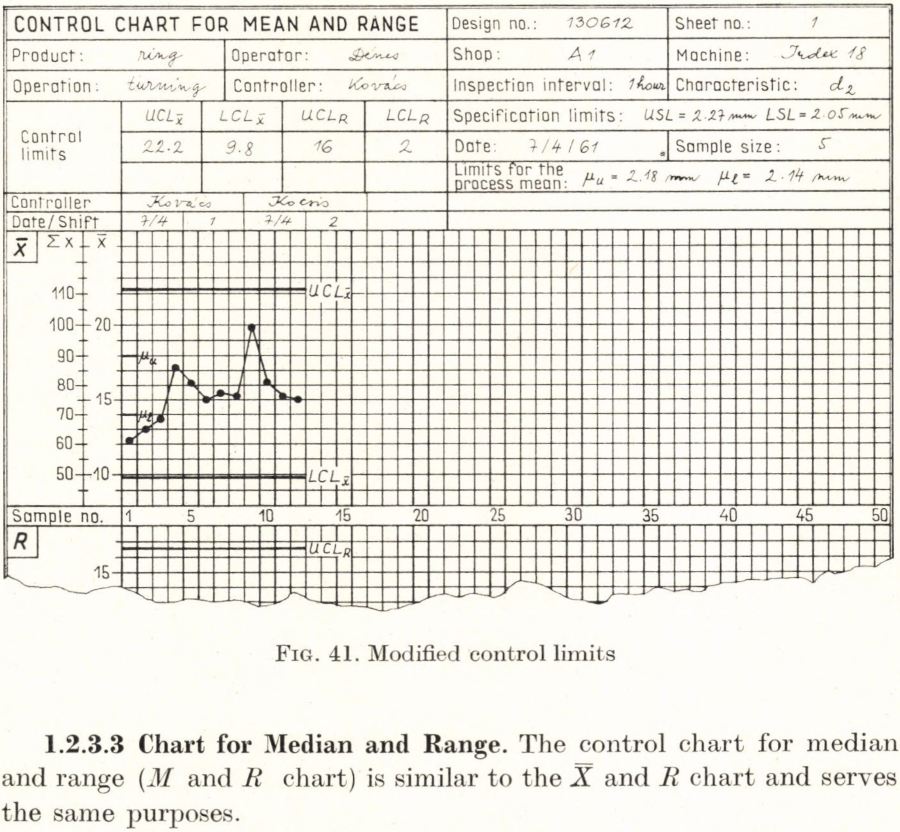
$$LCL_{\bar{X}} = \mu_l - A_2\bar{R}.$$

When using modified control limits, special attention should be given to the trend of the mean. If a definite shift towards either μ_l or μ_u is detected, the process is to be revised before the process mean reaches μ_l or μ_u even if no point has fallen outside the control limits. In this case it is particularly advisable to apply a range chart, in addition to the mean chart. Figure 41 represents the same data as Fig. 40, but with modified control limits.

Periodic review of the control limits. From time to time (usually after 40–50 samplings), the control limits and the central line are recalculated on the basis of data obtained from a certain number of samples preceding the recalculation.

Some modifications of the $\overline{X}$ and R chart will be discussed in Section 1.2.3.4.

An example is given at the end of the next section.



CONTROL CHART FOR MEAN AND RANGE

Design no.: 130612	Sheet no.: 1		
Product: ring	Operator: Dénes	Shop: A 1	Machine: Index 18
Operation: turning	Controller: Kovács	Inspection interval: 1 hour	Characteristic: d_2

Control limits

$UCL_{\bar{x}}$	$LCL_{\bar{x}}$	UCL_R	LCL_R
22.2	9.8	16	2

Specification limits: USL = 2.27 mm LSL = 2.05 mm

Date: 7/4/61

Sample size: 5

Limits for the process mean: μ_u = 2.18 mm μ_l = 2.14 mm

Controller	Kovács	Kocsis
Date/Shift	7/4 1	7/4 2

FIG. 41. Modified control limits

1.2.3.3 Chart for Median and Range. The control chart for median and range (M and R chart) is similar to the $\overline{X}$ and R chart and serves the same purposes.

In the median chart all measurements are represented by points, and the point of the median will be encircled if the number of observations is odd. In the case when the sample size is even, the two central points will be encircled. (It is advisable, however, to choose odd sample sizes if possible.) If any of the encircled points falls outside of the control limits, appropriate action must be taken to detect the source of the assignable change in the distribution.

The chart for median and range is applied by using the same basic principles as are used in the chart for mean and range. Some sources suggest the use of the same control limits as are used on the mean chart; this "approximation" is not justified, however.

The exact control limits of the median can be simply determined, not only for the normal case, but for any parent distribution. (See Chapter II, Section 3.3.1.) For the case of a normal parent distribution, Table 9 gives the constants F which can be used to calculate the median control limits UCL_M and LCL_M by the formulae

$$\mathrm{UCL}_M = \mu + F\bar{R}$$
$$\mathrm{LCL}_M = \mu - F\bar{R}.$$

Figure 42 represents the chart for median and range.

Example: Starting from the preliminary data given in the example of Section 1.2.2 let us determine the control limits of the chart for mean and range. Using Table 9, we obtain

$$\mathrm{UCL}_{\bar{X}} = \mu + A_2\bar{R} = 15 + 0.58 \cdot 7 = 19.2$$
$$\mathrm{LCL}_{\bar{X}} = \mu - A_2\bar{R} = 15 - 0.58 \cdot 7 = 10.8$$
$$\mathrm{UCL}_R = D_2\bar{R} = 2.34 \cdot 7 = 0.16$$
$$\mathrm{LCL}_R = D_1\bar{R} = 0.16 \cdot 7 = 0.02.$$

Since R is always an integer when expressed in terms of the smallest units to be read from the measuring instruments, we do not keep decimals in UCL_R and LCL_R. It is suitable to round off inward (i.e., round off the upper limit downwards and the lower limit upwards); on the other hand, we draw the control lines somewhat outward of the values obtained since these values are still acceptable (see Fig. 40).

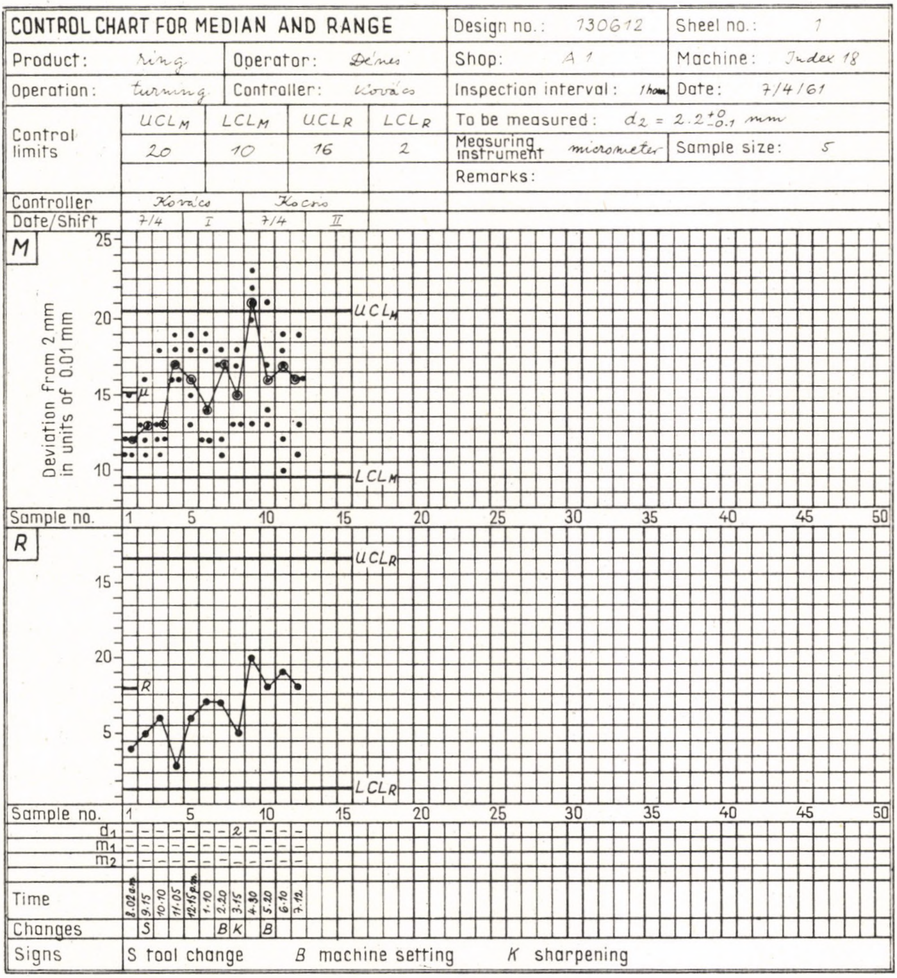
The control limits for the median, using the same initial data, are

$$\mathrm{UCL}_M = \mu + F\bar{R} = 15 + 0.72 \cdot 7 = 20$$
$$\mathrm{UCL}_M = \mu - F\bar{R} = 15 - 0.72 \cdot 7 = 10$$

(Fig. 42). In rounding off it is recommended to use the same principle as in the case of the chart for range.

Let us modify the original conditions so that the specification limits are, in the original units, 2.05 and 2.27, i.e., in our units 5 and 27. Let us use modified control limits on the chart for mean. The marginal values for the process mean are:

$$\mu_u = \mathrm{USL} - \frac{H}{2}\bar{R} = 27 - 1.29 \cdot 7 = 18$$
$$\mu_l = \mathrm{LSL} + \frac{H}{2}\bar{R} = 5 + 1.29 \cdot 7 = 14.$$



CONTROL CHART FOR MEDIAN AND RANGE

Design no.: 130612
Sheel no.: 1
Product: ring
Operator: Dénes
Shop: A 1
Machine: Index 18
Operation: turning
Controller: Kovács
Inspection interval: 1 hour
Date: 7/4/61

Control limits	UCL_M	LCL_M	UCL_R	LCL_R
	20	10	16	2

To be measured: $d_2 = 2.2^{+0}_{-0.1}$ mm
Measuring instrument: micrometer
Sample size: 5
Remarks:

Controller: Kovács | Kocris
Date/Shift: 7/4 I | 7/4 II

Sample no.	1	2	3	4	5	6	7	8	9	10	11	12
d_1	–	–	–	–	–	–	–	2	–	–	–	–
m_1	–	–	–	–	–	–	–	–	–	–	–	–
m_2	–	–	–	–	–	–	–	–	–	–	–	–
Time	8.02 a.m.	9.15	10.10	11.05	12.15 p.m.	1.10	2.20	3.15	4.30	5.20	6.10	7.12
Changes		S					B	K		B		

Signs: S tool change B machine setting K sharpening

M
25
20
15
10
Deviation from 2 mm in units of 0.01 mm
UCL_M
LCL_M
μ
Sample no.
1
5
10
15
20
25
30
35
40
45
50
R
15
20
5
UCL_R
LCL_R
$\bar{R}$

FIG. 42.

It is suitable to calculate these values before deciding on the machine setting. The modified control limits will be:

$$\mathrm{UCL}_{\overline{X}} = \mu_u + A_2\overline{R} = 18 + 0.58 \cdot 7 = 22.2$$

$$\mathrm{LCL}_{\overline{X}} = \mu_l - A_2\overline{R} = 14 - 0.58 \cdot 7 = 9.8.$$

(See Fig. 41.)

1.2.3.4 Some Other Types of Control Charts. Apart from the most frequently used types of control charts discussed in the last two sections

DATA SHEET
For control of mean and standard deviation

Sheet no.: *1/1*
Product: *Bicycle ball bearing*
Characteristic measured: *crumbling strength in q's*
Specification limits: *30 q, 60 q*

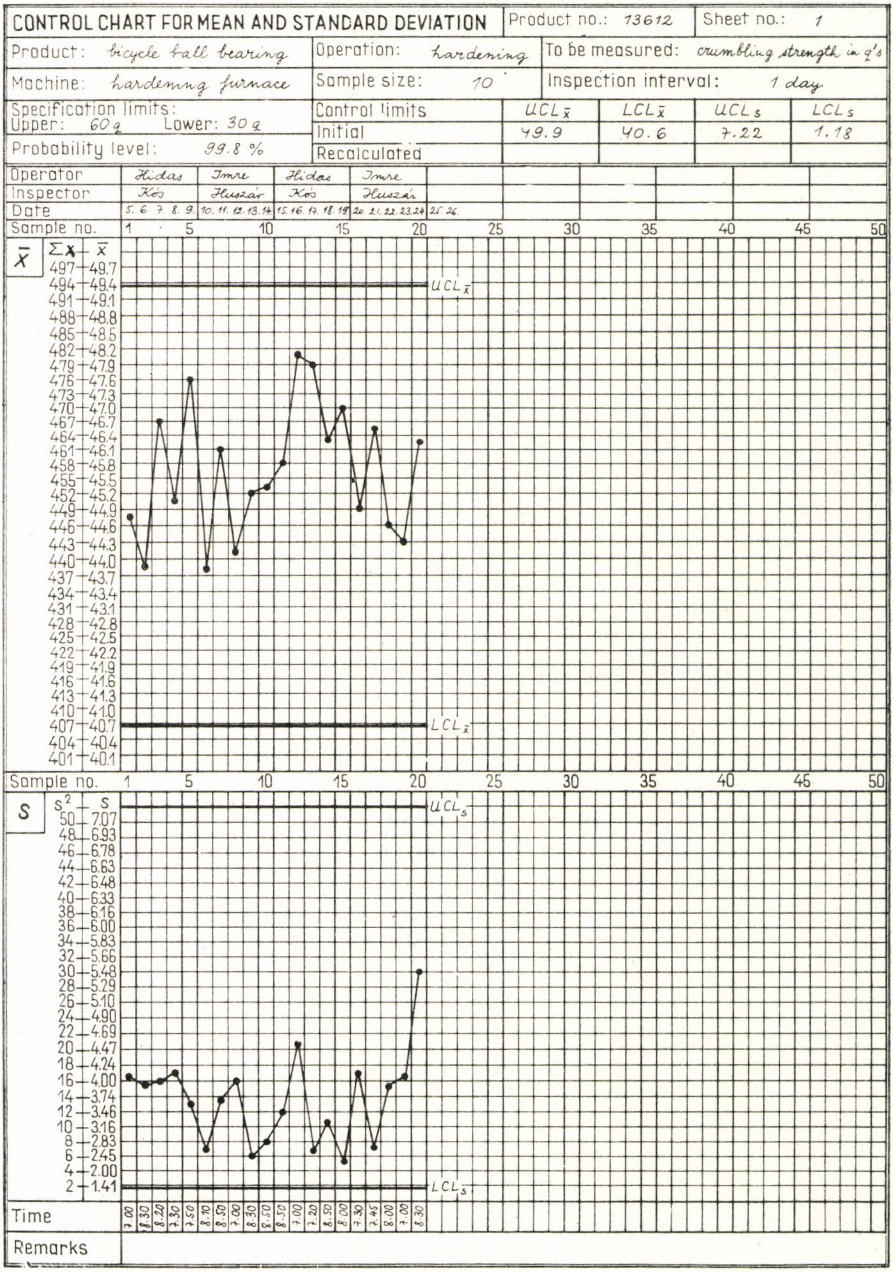
Machine: *hardening furnace*
Operator: *Hidas, Imre*
Controller: *Kós, Huszár*
Inspection interval: *1 day*
Sample size: *10*

		x	x^2		x	x^2		x	x^2		x	x^2
Sample no.	①	*48*	*2304*	②	*44*	*1936*	③	*51*	*2601*	④	*43*	*1849*
		43	*1849*		*40*	*1600*		*42*	*1764*		*46*	*2116*
Day:	*5*	*41*	*1681*	*6*	*40*	*1600*	*7*	*45*	*2025*	*8*	*35*	*1225*
		48	*2304*		*44*	*1936*		*50*	*2500*		*46*	*2116*
Time:	*7.00*	*39*	*1521*	*8.30*	*43*	*1849*	*8.20*	*46*	*2116*	*7.30*	*47*	*2209*
		49	*2401*		*41*	*1681*		*52*	*2704*		*43*	*1849*
Series no.	*583*	*41*	*1681*	*584*	*50*	*2500*	*585*	*44*	*1936*	*586*	*49*	*2401*
		45	*2025*		*39*	*1521*		*49*	*2401*		*51*	*2601*
		52	*2704*		*48*	*2304*		*49*	*2401*		*46*	*2116*
		42	*1764*		*50*	*2500*		*39*	*1521*		*45*	*2025*
Sum:		*448*	*20234*		*439*	*19427*		*467*	*21969*		*451*	*20507*
Mean: ($\bar{x}$, $\overline{x^2}$)		*44.8*	*2023.4*		*43.9*	*1942.7*		*46.7*	*2196.9*		*45.1*	*2050.7*
Square of mean: ($\bar{x}^2$)			*2007.04*			*1927.21*			*2180.89*			*2034.01*
Variance: (s^2)			*16.36*			*15.49*			*16.01*			*16.69*

Sample no.:	⑤	50	2500	⑥	44	1936	⑦	42	1764	⑧	45	2025
		44	1936		44	1936		52	2704		36	1296
Day:	9	50	2500	10	48	2304	11	45	2025	12	40	1600
		52	2704		42	1764		49	2401		48	2304
Time:	7.50	51	2601	8.10	46	2116	8.50	45	2025	7.00	42	1764
		51	2601		38	1444		48	2304		50	2500
Series no.	587	42	1764	588	46	2116	589	43	1849	590	42	1764
		46	2116		43	1849		52	2704		48	2304
		42	1764		45	2025		43	1849		44	1936
		48	2304		42	1764		42	1764		46	2116
Sum:		476	22790		438	19254		461	21389		441	19609
Mean: ($\bar{x}$, $\overline{x^2}$)		47.6	2279.0		43.8	1925.4		46.1	2138.9		44.1	1960.9
Square of mean: ($\bar{x}^2$)			2265.76			1918.44			2125.21			1944.81
Variance: (s^2)			13.24			6.96			13.69			16.09
Remark:												

FIG. 43. Data sheet for $\bar{X}$ & *R* chart

(Filling in is exemplified by *Italics*)



CONTROL CHART FOR MEAN AND STANDARD DEVIATION		Product no.: 13612		Sheet no.: 1	
Product: bicycle ball bearing	Operation: hardening		To be measured: crumbling strength in g's		
Machine: hardening furnace	Sample size: 10		Inspection interval: 1 day		
Specification limits: Upper: 60 g Lower: 30 g	Control limits	$UCL_{\bar{x}}$	$LCL_{\bar{x}}$	UCL_s	LCL_s
	Initial	49.9	40.6	7.22	1.18
Probability level: 99.8 %	Recalculated				

Operator	Hidas	Imre	Hidas	Imre	
Inspector	Kós	Huszár	Kós	Huszár	
Date	5. 6. 7. 8. 9.	10. 11. 12. 13. 14.	15. 16. 17. 18. 19.	20. 21. 22. 23. 24.	25. 26.

Time: 7.00, 8.30, 8.20, 7.30, 7.50, 8.10, 8.50, 7.00, 8.30, 8.50, 8.50, 7.00, 7.20, 8.50, 8.00, 7.30, 7.45, 8.00, 7.00, 8.30

Remarks

Sample no.
1 5 10 15 20 25 30 35 40 45 50
$\bar{x}$
Σx — $\bar{x}$
497 — 49.7
494 — 49.4
491 — 49.1
488 — 48.8
485 — 48.5
482 — 48.2
479 — 47.9
476 — 47.6
473 — 47.3
470 — 47.0
467 — 46.7
464 — 46.4
461 — 46.1
458 — 45.8
455 — 45.5
452 — 45.2
449 — 44.9
446 — 44.6
443 — 44.3
440 — 44.0
437 — 43.7
434 — 43.4
431 — 43.1
428 — 42.8
425 — 42.5
422 — 42.2
419 — 41.9
416 — 41.6
413 — 41.3
410 — 41.0
407 — 40.7
404 — 40.4
401 — 40.1
$UCL_{\bar{x}}$
$LCL_{\bar{x}}$
s
s^2 — s
50 — 7.07
48 — 6.93
46 — 6.78
44 — 6.63
42 — 6.48
40 — 6.33
38 — 6.16
36 — 6.00
34 — 5.83
32 — 5.66
30 — 5.48
28 — 5.29
26 — 5.10
24 — 4.90
22 — 4.69
20 — 4.47
18 — 4.24
16 — 4.00
14 — 3.74
12 — 3.46
10 — 3.16
8 — 2.83
6 — 2.45
4 — 2.00
2 — 1.41
UCL_s
LCL_s

FIG. 44.

many special kinds of control charts or techniques may prove to be helpful in certain applications. Here we mention the following:

(a) *Chart for Standard Deviation.* This type of chart may be applied instead of the R chart whenever the use of all the information in the sample is more important than simplicity of calculation. The upper control limit is obtained by multiplying the estimate of the standard deviation by B_2 from Table 9. The population standard deviation should be estimated by the square root of the average of the subgroup variances rather than the average of the subgroup standard deviations. The other method is inefficient and furthermore requires more computation.

The calculation of the square root may be avoided in two ways: apply a *chart for variance* or *sum of squares* instead of the chart for standard deviation, or furnish the chart with a double scale. The former procedure may cause some difficulties in the practical interpretation of the chart, whereas the latter may increase the possibility of errors arising from confusion.

Figures 43 and 44 represent the chart for standard deviation and its data sheet, respectively.

(b) *Chart for order statistics.* The chart for order statistics is the generalization of the chart for median, discussed in Section 1.2.3.3. The control limits of the kth ordered element of a sample are calculated by the formulae

$$\mathrm{UCL}_{(k)} = \mu + F_{(k)}\,\bar{R},$$

$$\mathrm{LCL}_{(k)} = \mu - A_{(k)}\,\bar{R},$$

the constants $A_{(k)}$ and $F_{(k)}$ being given, for normal distribution, in Table 10. (Concerning the basic theory we refer to II. 3.3.1.)

For other underlying distributions, the limits can be easily calculated with the help of column (1). For other probability levels the table of incomplete beta function can be used [C 9]. Cf. also [B 50, B 27].

In the chart represented in Fig. 45, three of the ordered elements are controlled.

Another important special case of the chart for order statistics is the *chart for the largest* and *the smallest value* (Fig. 46), [B 42]).

(c) *Chart for midrange.* The chart for midrange can be used for the same purpose as the chart for mean or chart for median. It saves calculation work as does the median chart.

For coefficients and details see [A 7], pp. 245—6.

(d) *Sloping control lines.* If the shift in the process mean or dispersion increase is regular, the use of sloping control lines may be considered.

TABLE 10

CONTROL LIMITS OF ORDER STATISTICS

(1) Values of $F(x_1)$ and $F(x_2)$, where $F(x)$: distribution function of (continuous) parent distribution; $F_{nk}(x)$: distribution function of the kth out of n ordered elements; (2) x_1 and x_2 for standardized normal parent distribution; (3) $A_{(k)} = x_1/d_2$ and $F_{(k)} = x_2/d_2$; where x_1 and x_2: entries in column (2); d_2: expectation of sample range for standardized normal distribution; $A_{(k)}\overline{R}$ and $F_{(k)}\overline{R}$: distance of control limits of the kth ordered element from the mean of the parent distribution ($\overline{R}$: mean range)

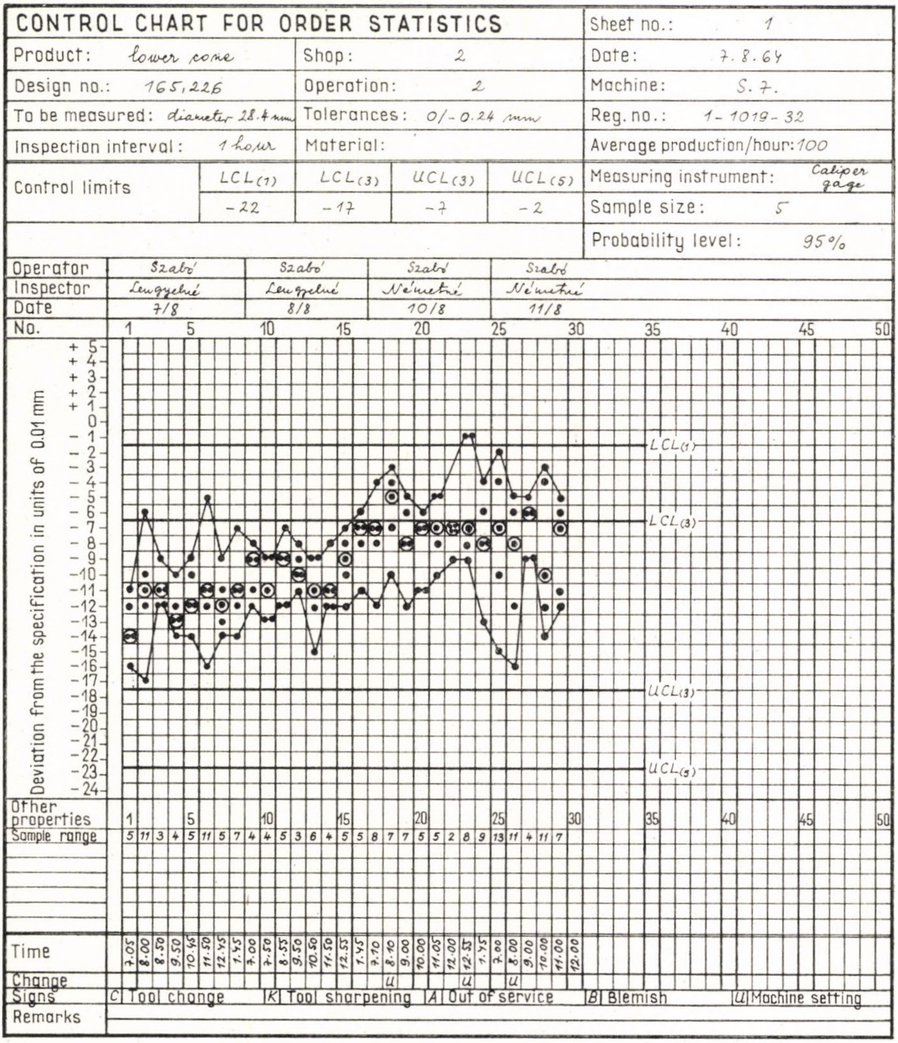
95%								99%							
Lower limit ($F_{nk}(x_1)=0.025$)				Upper limit ($F_{nk}(x_2)=0.975$)				Lower limit ($F_{nk}(x_1)=0.005$)				Upper limit ($F_{nk}(x_2)=0.995$)			
k	(1)	(2)	(3)	k	(1)	(2)	(3)	k	(1)	(2)	(3)	k	(1)	(2)	(3)
$n = 4$								$n = 4$							
1	0.006	−2.49	−1.21	1	0.602	+0.26	+0.13	1	0.001	−3.02	−1.47	1	0.734	+0.63	+0.31
2	0.068	−1.49	−0.72	2	0.806	+0.86	+0.42	2	0.030	−1.88	−0.91	2	0.889	+1.22	+0.60
3	0.194	−0.86	−0.42	3	0.932	+1.49	+0.72	3	0.111	−1.22	−0.59	3	0.970	+1.88	+0.91
4	0.398	−0.26	−0.13	4	0.994	+2.49	+1.21	4	0.266	−0.63	-0.31	4	0.999	+3.02	+1.47
$n = 5$								$n = 5$							
1	0.005	−2.57	−1.11	1	0.522	+0.05	+0.02	1	0.001	−3.09	−1.33	1	0.653	+0.39	+0.17
2	0.052	−1.63	−0.70	2	0.717	+0.57	+0.25	2	0.025	−1.96	−0.84	2	0.815	+0.90	+0.39
3	0.146	−1.05	−0.45	3	0.854	+1.05	+0.45	3	0.083	−1.39	−0.60	3	0.917	+1.39	+0.60
4	0.283	−0.57	−0.25	4	0.948	+1.63	+0.70	4	0.185	−0.90	−0.39	4	0.975	+1.96	+0.84
5	0.478	−0.05	−0.02	5	0.995	+2.57	+1.11	5	0.347	−0.39	−0.17	5	0.999	+3.09	+1.33

n = 6							
1	0.004	—2.63	—1.04	1	0.459	—0.10	—0.04
2	0.043	—1.72	—0.68	2	0.641	+0.36	+0.14
3	0.118	—1.19	—0.47	3	0.777	+0.76	+0.30
4	0.223	—0.76	—0.30	4	0.882	+1.19	+0.47
5	0.359	—0.36	—0.14	5	0.957	+1.72	+0.68
6	0.541	+0.10	+0.04	6	0.996	+2.63	+1.04
n = 7							
1	0.004	—2.69	—1.00	1	0.410	—0.23	—0.09
2	0.037	—1.79	—0.66	2	0.581	+0.20	+0.07
3	0.099	—1.29	—0.48	3	0.710	+0.55	+0.20
4	0.184	—0.90	—0.33	4	0.816	+0.90	+0.33
5	0.290	—0.55	—0.20	5	0.901	+1.29	+0.48
6	0.419	—0.20	—0.07	6	0.963	+1.79	+0.66
7	0.590	+0.23	+0.09	7	0.996	+2.69	+1.00
n = 8							
1	0.003	—2.73	—0.96	1	0.369	—0.33	—0.12
2	0.032	—1.85	—0.65	2	0.527	+0.07	+0.02
3	0.085	—1.37	—0.48	3	0.651	+0.39	+0.14
4	0.154	—1.02	—0.36	4	0.755	+0.69	+0.24
5	0.254	—0.69	—0.24	5	0.846	+1.02	+0.36
6	0.349	—0.39	—0.14	6	0.915	+1.37	+0.48
7	0.473	—0.07	—0.02	7	0.968	+1.85	+0.65
8	0.631	+0.33	+0.12	8	0.997	+2.73	+0.96

n = 6							
1	0.001	—3.14	—1.24	1	0.586	+0.22	+0.09
2	0.019	—2.08	—0.82	2	0.745	+0.66	+0.26
3	0.066	—1.51	—0.60	3	0.855	+1.06	+0.42
4	0.145	—1.06	—0.42	4	0.934	+1.51	+0.60
5	0.255	—0.66	—0.26	5	0.981	+2.08	+0.82
6	0.414	—0.22	—0.09	6	0.999	+3.14	+1.24
n = 7							
1	0.001	—3.19	—1.18	1	0.531	+0.08	+0.03
2	0.016	—2.14	—0.79	2	0.689	+0.48	+0.18
3	0.055	—1.60	—0.59	3	0.979	+0.83	+0.31
4	0.118	—1.19	—0.44	4	0.882	+1.19	+0.44
5	0.203	—0.83	—0.31	5	0.945	+1.60	+0.59
6	0.315	—0.48	—0.18	6	0.984	+2.14	+0.79
7	0.469	—0.08	—0.03	7	0.999	+3.49	+1.18
n = 8							
1	0.001	—3.23	—1.14	1	0.484	—0.04	—0.01
2	0.014	—1.89	—0.66	2	0.632	+0.34	+0.12
3	0.047	—1.68	—0.59	3	0.742	+0.65	+0.23
4	0.100	—1.28	—0.45	4	0.830	+0.95	+0.34
5	0.170	—0.95	—0.34	5	0.900	+1.28	+0.45
6	0.258	—0.65	—0.23	6	0.953	+1.68	+0.59
7	0.368	—0.34	—0.12	7	0.986	+1.89	+0.66
8	0.516	+0.04	+0.01	8	0.999	+3.23	+1.14

TABLE 10 (cont.)

95%								99%							
Lower limit ($F_{nk}(x_1)=0.025$)				Upper limit ($F_{nk}(x_2)=0.975$)				Lower limit ($F_{nk}(x_1)=0.005$)				Upper limit ($F_{nk}(x_2)=0.995$)			
k	(1)	(2)	(3)	k	(1)	(2)	(3)	k	(1)	(2)	(3)	k	(1)	(2)	(3)
$n = 9$								$n = 9$							
1	0.003	−2.77	−0.93	1	0.336	−0.42	−0.14	1	0.002	−3.26	−1.10	1	0.445	−0.14	−0.05
2	0.028	−1.91	−0.64	2	0.483	−0.04	−0.01	2	0.012	−2.27	−0.76	2	0.585	+0.21	+0.07
3	0.075	−1.44	−0.49	3	0.600	+0.25	+0.08	3	0.041	−1.74	−0.59	3	0.693	+0.50	+0.17
4	0.133	−1.11	−0.37	4	0.701	+0.52	+0.18	4	0.096	−1.37	−0.46	4	0.781	+0.78	+0.26
5	0.212	−0.80	−0.27	5	0.788	+0.80	+0.27	5	0.146	−1.05	−0.35	5	0.854	+1.05	+0.35
6	0.299	−0.52	−0.18	6	0.867	+1.11	+0.37	6	0.219	−0.78	−0.26	6	0.914	+1.37	+0.46
7	0.400	−0.25	−0.08	7	0.925	+1.44	+0.49	7	0.307	−0.50	−0.17	7	0.959	+1.74	+0.59
8	0.517	+0.04	+0.01	8	0.972	+1.91	+0.64	8	0.415	−0.21	−0.07	8	0.988	+2.27	+0.76
9	0.664	+0.42	+0.14	9	0.997	+2.77	+0.93	9	0.555	+0.14	+0.05	9	0.998	+3.26	+1.10
$n = 10$								$n = 10$							
1	0.003	−2.80	−0.91	1	0.308	−0.50	−0.16	1	0.001	−3.29	−1.07	1	0.411	−0.23	−0.07
2	0.025	−1.96	−0.64	2	0.445	−0.14	−0.05	2	0.011	−2.30	−0.75	2	0.544	+0.11	+0.04
3	0.066	−1.51	−0.49	3	0.556	+0.14	+0.05	3	0.036	−1.80	−0.58	3	0.648	+0.38	+0.12
4	0.121	−1.17	−0.38	4	0.652	+0.39	+0.13	4	0.076	−1.43	−0.46	4	0.735	+0.63	+0.20
5	0.187	−0.89	−0.29	5	0.738	+0.64	+0.21	5	0.128	−1.14	−0.37	5	0.809	+0.87	+0.28
6	0.262	−0.64	−0.21	6	0.813	+0.89	+0.29	6	0.191	−0.87	−0.28	6	0.872	+1.14	+0.37
7	0.348	−0.39	−0.13	7	0.879	+1.17	+0.38	7	0.265	−0.63	−0.20	7	0.924	+1.43	+0.46
8	0.444	−0.14	−0.05	8	0.934	+1.51	+0.49	8	0.352	−0.38	−0.12	8	0.964	+1.80	+0.58
9	0.555	+0.14	+0.05	9	0.975	+1.96	+0.64	9	0.456	−0.11	−0.04	9	0.989	+2.30	+0.75
10	0.692	+0.50	+0.16	10	0.997	+2.80	+0.91	10	0.589	+0.23	+0.07	10	0.999	+3.29	+1.07



CONTROL CHART FOR ORDER STATISTICS

Product: lower cone	Shop: 2	Sheet no.: 1
Design no.: 165,226	Operation: 2	Date: 7.8.64
To be measured: diameter 28.4 mm	Tolerances: 0/-0.24 mm	Machine: S.7.
Inspection interval: 1 hour	Material:	Reg. no.: 1-1019-32
		Average production/hour: 100

Control limits	LCL(1)	LCL(3)	UCL(3)	UCL(5)	Measuring instrument: Caliper gage
	-22	-17	-7	-2	Sample size: 5
					Probability level: 95%

Operator	Szabó	Szabó	Szabó	Szabó
Inspector	Lengyelné	Lengyelné	Németné	Németné
Date	7/8	8/8	10/8	11/8

No.
1
5
10
15
20
25
30
35
40
45
50
Deviation from the specification in units of 0.01 mm
+5
+4
+3
+2
+1
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
-12
-13
-14
-15
-16
-17
-18
-19
-20
-21
-22
-23
-24
LCL(1)
LCL(3)
UCL(3)
UCL(5)
Other properties

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sample range	5	11	3	4	5	11	5	7	4	4	5	3	6	4	5
Time	7.05	8.00	8.50	9.50	10.45	11.50	12.45	1.45	7.00	7.50	8.55	9.50	10.50	11.50	12.55
Change															

No.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Sample range	5	8	7	7	5	5	2	8	9	13	11	4	11	7	
Time	1.45	7.10	8.10	9.00	10.00	11.05	12.00	12.55	1.45	7.00	8.00	9.00	10.00	11.00	12.00
Change			U					U			U				

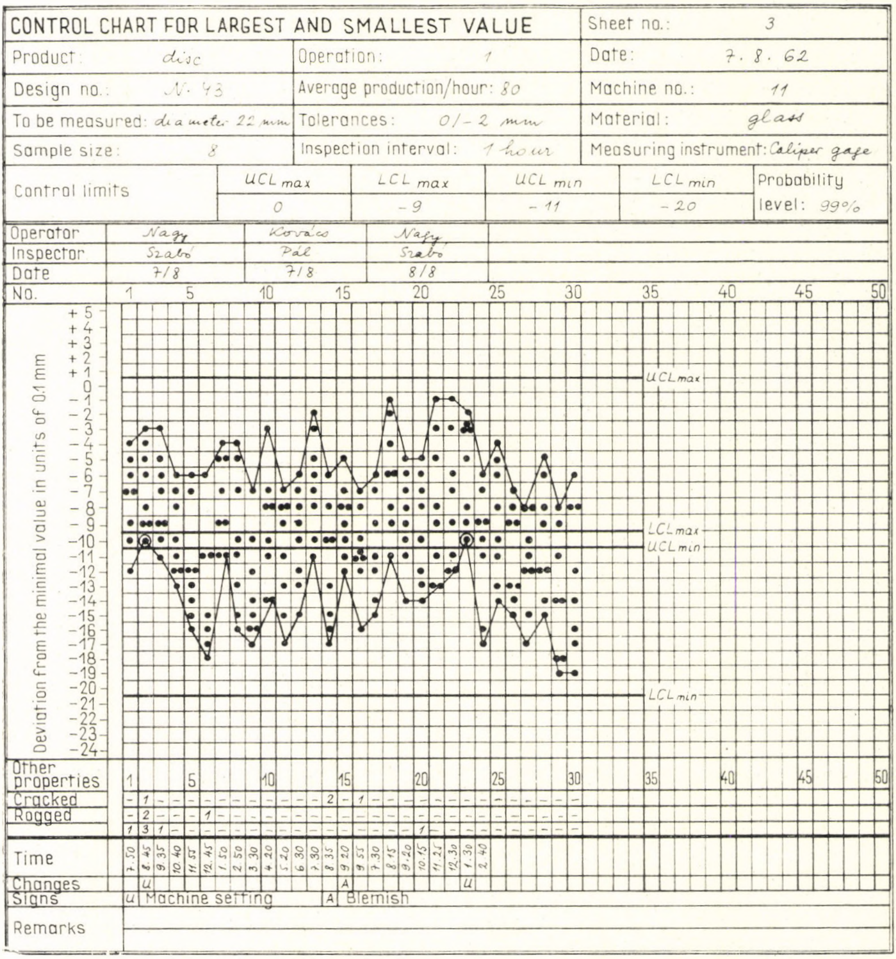
Signs: C Tool change; K Tool sharpening; A Out of service; B Blemish; U Machine setting

Remarks

FIG. 45.

In this case a trend term is to be added to the formulas, and/or the coefficient of R or σ is to be replaced by a function expressing the time variation. Cf. also, e.g., [A 21], pp. 121–3.

(e) *Warning limits and runs.* Some authors dealing with statistical quality control suggest the use of two sets of limits on the control charts. The inner limits are called *warning limits.* They are usually (approxi-



CONTROL CHART FOR LARGEST AND SMALLEST VALUE

CONTROL CHART FOR LARGEST AND SMALLEST VALUE		Sheet no.: 3
Product: disc	Operation: 1	Date: 7. 8. 62
Design no.: N. 43	Average production/hour: 80	Machine no.: 11
To be measured: diameter 22 mm	Tolerances: 0/−2 mm	Material: glass
Sample size: 8	Inspection interval: 1 hour	Measuring instrument: Caliper gage

Control limits	UCL max	LCL max	UCL min	LCL min	Probability level: 99%
	0	−9	−11	−20	

Operator	Nagy	Kovács	Nagy
Inspector	Szabó	Pál	Szabó
Date	7/8	7/8	8/8

Other properties	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Cracked	–	1	–	–	–	–	–	–	–	–	–	–	–	2	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Rogged	–	2	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	1	3	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–
Time	7.50	8.45	9.35	10.40	11.55	12.45	1.50	2.50	3.30	4.20	5.20	6.30	7.30	8.35	9.20	9.55	7.30	8.15	9.20	10.15	11.25	12.30	1.30	2.40						
Changes		U													A								U							

Signs: U Machine setting A Blemish

Remarks

FIG. 46.

mately) 95% probability limits. (Their distance from the central line is about two thirds of that from the control limits.)

If the warning limits are exceeded, the suspicion for the presence of an assignable change is reasonable. Also if two (or more) consecutive points exceed the same warning limit, it is reasonable to conclude that the process is out of control. (In fact, $0.025^2 = 0.00063$ is less than the unilateral probability of exceeding the control limits.) More generally, rules have been worked out on the basis of the binomial distribution: the minimal number of points have been determined that must exceed the warning limit in a run of given length for bringing about the same conclusion. If such rules are used in combination with the use of control limits it is advisable to compute the joint probability of the error of first kind.

Similar rules have been worked out for runs of points above (below) the central lines.

These versions utilize the information lying in the total configuration of the points in the chart. Therefore, they increase the sensitivity of the control chart to shift, if applied on the $\bar{X}$ chart or M chart, as hinted in Section 1.2.3.2.

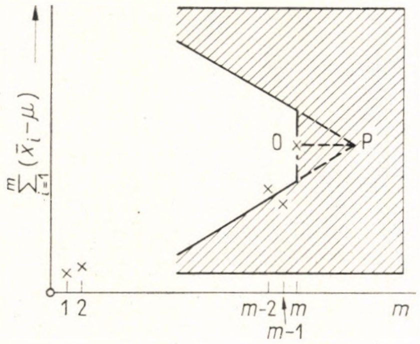


FIG. 47. Cumulative sum control chart

(f) *Cumulative sum chart.* The foregoing idea reaches its full development in the *cumulative sum chart.* Here each observation (or sample mean) is diminished by the process mean μ, and the cumulative sums of the deviates are represented in a coordinate system with the horizontal line being the time axis. Instead of control limits, a pattern is placed on the chart, its shape being represented by the shadowed area of Fig. 47. After each sampling this pattern has to be shifted in such a way that the point O will be in the last plotted point of the chart, the vector $\overrightarrow{OP}$ directed horizontally. If any of the plotted points is covered by the pattern, this is an indication of an assignable change.

This method was first suggested in its basic form by Page [57]. For more details, see Johnson and Leone [A 26], pp. 320 – 329.

(g) *Modified control limits and acceptance control chart.* A type of the modified control limit method was discussed in Section 1.2.3.2. It was mentioned there that in the case of a trend in the process mean, there is usually small probability that the control chart will give a sign as soon as the process mean exceeds its admissible marginal level. The parallel observation of the configuration of points was recommended. Alternatively the probability level may be decreased.

A completely different principle is that of determining the control limits within the extremal admissible values of the process mean (see, e.g., Juran [A 27], p. 13/63—65. It is quite natural that this method sets higher requirements for the width of the tolerance than the former one. The limits are determined to make practically certain that this control chart will give a signal before the process mean exceeds the extremal admissible value. On the other hand, the probability of the error of the first kind increases in this case.

These two aspects are united within the method of the so-called *acceptance control chart* (R. A. Freund [B 28, B 29]). Here the zones of the process mean are limited by two pairs of levels: the outer ones are called rejectable process levels and the inner ones acceptable process levels.

Let α be the probability that a process centered at an acceptable process level will be rejected, and β the probability that a process centered at a rejectable process level will be accepted. If α and β are predetermined, the sample size and the control limits can be calculated. If $\alpha = \beta$ and the underlying distribution is symmetric, the control limits will lie halfway between the acceptable process level and the rejectable process level lines. The similarity between this method and that of constructing sampling plans on the basis of two points of the OC curve is obvious (cf. Section 2.2.2); hence the name.

(h) *Multivariate charts.* If the control chart method is used simultaneously for two or more quality characteristics of the same product, the joint probability of the error of the first kind can be calculated; e.g. if the characteristics are independent, with the multiplication rule (see, e.g., Johnson and Leone [A 26], pp. 347—348). This unified chart, known as the multivariate chart, is not to be confused with the "multi-vari chart" of Juran and Seder ([A 27], pp. 11/30—32).

(i) The list of control charts and techniques based on the control of variables is by no means exhaustively discussed in this book. Control charts may be constructed for any sample statistic. The control charts can be used advantageously wherever a statistical test is repeatedly performed, not only in process control but in acceptance sampling, control of clerical errors, etc. It is a short-cut method for variance analysis [B 17].

1.3 CONTROL BY ATTRIBUTES

The circumstances for using control chart by attributes were specified in Section 1.1.2. It is evident that in spite of the advantage given by charts for variables, there are many cases in which only the attribute method can be applied or in which it is most suitable.

The observed variable here is either the number of defective units or the number of defects appearing in the sample. If the defective items (defects) occur independently, then the binomial (the Poisson) distribution is adequate (II. 2.9.3 and 2.9.2). Since in practice the fraction defective is usually below 10%, in the majority of cases the Poisson approximation can be used for the binomial case as well.

If we check the proportion of defective items, then the *chart for number of defectives* (*np chart*, Section 1.3.2) or the *chart for fraction defective* (*p chart*, Section 1.3.3.) can be used. For constant sample size the former is more advantageous, since it saves one calculation: the division by the sample size. If the sample size shows considerable fluctuation, then only the p chart can be used.

If we wish to control the number of defects, the *chart for defects* (*c chart*, Section 1.3.5) should be used in the case of a constant sample size, otherwise we use the *chart for defects per unit* (Section 1.3.5). These charts can be used not only in case of piece goods but for continuous products (e.g., yarn, roll of paper, etc.) as well.

The computation is based on the average fraction defective or average number of defects. If this is not known by past experience, a preliminary sampling (*process analysis*) should be performed for this purpose (Sections 1.3.1 and 1.3.4).

1.3.1 Process Analysis for Defectives

The estimation of the fraction defective is a very simple problem: the number of defective items is to be divided by the number of all items investigated.

It is, however, desirable to have a reliable estimate for computing control limits.

This requirement implies a rather high amount of basic data. If we wish to have an estimate with 95% confidence limits which do not differ from the actual value of the fraction defective by more than 10%, then the preliminary data must contain about 400 defective items. For example, if the fraction defective is near 1%, the inspection of 40,000 items is necessary in order to conclude, at a 95% confidence level, that the estimate is between such narrow limits as 0.9% and 1.1%. This requirement is rather high for products with a very small fraction defective.

For methods assuring homogeneity and representativeness in the preliminary sampling we refer to the relevant remarks made in Section 1.2.1.1.

It is desirable to test the preliminary data to see if it can be regarded as independent repetitions of the same experiment. For this purpose the χ^2 test is appropriate (see II. 4.3.7.6). However, under practical circumstances, it is usually more suitable to form subgroups in chronological order and to compare the number of defectives in each subgroup with the control limits (for their computation see the next section).

1.3.2 Chart for the Number of Defectives

In calculating significance limits, we may use tables of the binomial distribution, or use Poisson or normal approximation.

As mentioned before, in almost all cases of process control the binomial distribution can be well approximated by the Poisson distribution. This fact makes it possible to construct simple tables for the calculation of the control limits of the *np* chart.

In Table 11 the upper and lower probability limits UCL_c and LCL_c are given, as a function of the average number of defective items $c = np$. (n denotes the sample size, and p the fraction defective determined from past data.) The calculations of this table have been based on the probability level 99.8%, with equal probability of exceeding the limits in both directions. Owing to the discrete character of the Poisson distribution, these values are only approximate. In the elaboration of the table we deviated slightly from the convention: instead of considering the significance level as an upper bound of the actual probability of the error of the first kind, we chose this probability as close as possible to its prescribed level. This seemed to be a more appropriate principle since the level is basically a matter of arbitrary choice.

Example: if $c = 7.8$, $\text{UCL}_c = 17$, $\text{LCL}_c = 1$. If the value of c appears as a limiting value in the table, either of the two values of UCL_c (LCL_c) can be chosen. (The probability that the smaller value of UCL_c will be exceeded is as far above .001 as the probability of exceeding the larger value is below .001: e.g., if $c = 7.9$ and $\text{UCL}_c = 17$ are chosen, the actual error of the first kind is 1.4‰, and it is 0.6‰ if $\text{UCL}_c = 18$.)

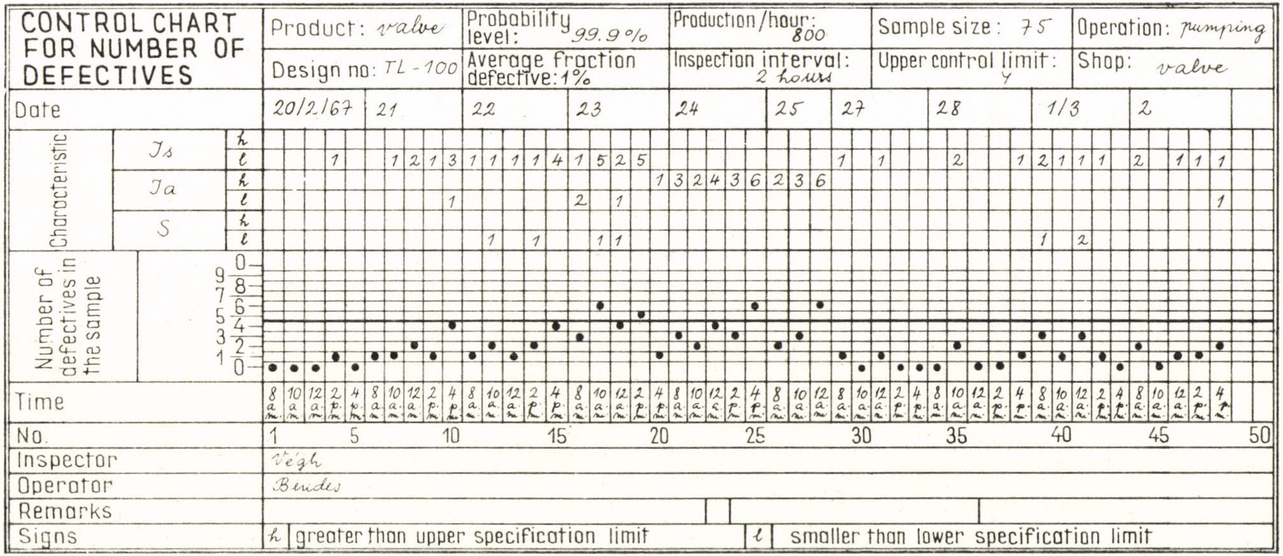
TABLE 11

CONTROL LIMITS FOR np CHARTS AND c CHARTS

Probability level: 99.8% (0.1% probability on either side)

c	LCL$_c$	c	UCL$_c$	Nominal value of c
		0.0005 … 0.002	0	0.001
		0.002 … 0.065	1	0.045
		0.065 … 0.23	2	0.19
		0.23 … 0.52	3	0.45
		0.52 … 0.84	4	0.74
		0.84 … 1.2	5	1.1
		1.2 … 1.7	6	1.5
		1.7 … 2.1	7	2.0
		2.1 … 2.6	8	2.5
		2.6 … 3.1	9	3.0
		3.1 … 3.7	10	3.5
		3.7 … 4.2	11	4.0
		4.2 … 4.8	12	4.6
		4.8 … 5.4	13	5.2
		5.4 … 6.0	14	5.8
0 … 6.2	0	6.0 … 6.6	15	6.4
		6.6 … 7.3	16	7.0
		7.3 … 7.9	17	7.7
6.2 … 8.6	1	7.9 … 8.5	18	8.3
		8.5 … 9.2	19	9.0
		9.2 … 9.9	20	9.6
8.6 … 10.6	2	9.9 … 10.6	21	10.3
		10.6 … 11.2	22	11.0
10.6 … 12.4	3	11.2 … 11.9	23	11.7
		11.9 … 12.6	24	12.4
		12.6 … 13.3	25	13.1
12.4 … 14.2	4	13.3 … 14.0	26	13.8
		14.0 … 14.7	27	14.5
14.2 … 15.9	5	14.7 … 15.4	28	15.2
		15.4 … 16.1	29	15.8
		16.1 … 16.9	30	16.6
15.9 … 17.5	6	16.9 … 17.5	31	17.3
17.5 … 19.0	7	17.5 … 18.3	32	18.0
		18.3 … 19.1	33	18.8
		19.1 … 19.8	34	19.5
19.0 … 20.6	8	19.8 … 20.5	35	20.2

Note. The probability that a point falls outside the upper (or outside the lower) control limit is as near to 0.1% as possible. It is equal to 0.1% for the upper control limit if the expected value of the control chart variable (c) takes on the "nominal value" (last column).



CONTROL CHART FOR NUMBER OF DEFECTIVES

Product: valve
Probability level: 99.9 %
Production/hour: 800
Sample size: 75
Operation: pumping
Design no: TL-100
Average fraction defective: 1%
Inspection interval: 2 hours
Upper control limit: 4
Shop: valve

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Date	20/2/67					21					22					23					24			
Characteristic Ts h																								
Ts l				1			1	2	1	3	1	1	1	1	4	1	5	2	5					
Ta h																				1	3	2	4	3
Ta l										1						2		1						
S h																								
S l												1		1			1	1						
Time	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.
No.	1				5					10					15					20				

	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Date		25			27					28					1/3					2						
Characteristic Ts h																										
Ts l					1		1				2			1	2	1	1	1		2		1	1	1		
Ta h	6	2	3	6																						
Ta l																								1		
S h																										
S l															1		2									
Time	4 p.m.	8 a.m.	10 a.m.	12 a.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.	8 a.m.	10 a.m.	12 a.m.	2 p.m.	4 p.m.		
No.	25					30					35					40					45					50

Number of defectives in the sample

Inspector: Végh
Operator: Bendes
Remarks

Signs: h greater than upper specification limit; l smaller than lower specification limit

Fig. 48.

Although the tables give values for small values of c as well, it is suitable to choose a sample size such that UCL_c would be at least a few units.

The lower limits are not always used. (If not, then the probability limit is but 99.9%.)

Figure 48 shows the use of the control chart for the number of defectives.

If the sample size shows small fluctuations we may still use this type of control chart. (In the case when the corresponding fluctuation of the control limit does not exceed unity, its use may be considered admissible.) In this case the control limits are to be calculated according to the average sample size.

In most other respects we may apply the instructions given for control charts for variables.

1.3.3 Control Chart for Fraction Defective

In many cases there is considerable fluctuation in the sample size, so that the chart described in the preceding Section cannot be applied. Instead the *control chart for fraction defective* is appropriate.

Here new control limits should, in principle, be calculated for each subgroup. If, however, the fluctuation in the sample size is moderate, fixed limits can be computed based on the average sample size. (Although this simplification distorts the probability level, the deviation is less than if the control chart for the number of defectives is used.) The formulas for the control limits, UCL_p and LCL_p, are as follows:

$$UCL_p = \frac{UCL_c + 0.5}{n},$$

$$LCL_p = \frac{LCL_c - 0.5}{n},$$

where UCL_c and LCL_c are determined by the method described in the preceding Section and n is the sample size (average sample size). The corrections ±0.5 are for continuity since here both the subgroup fraction defectives and the control limits are plotted on the same continuous scale. (If exact sample sizes are used in the calculation, the correction does not change the results but makes their reading easier. If average sample sizes are used, the correction improves the approximation.)

1.3.4 Narrow-Limit Gaging

As already mentioned in 1.1.2, if a gage is used for the control of a measurable characteristic, the method of attributes — i.e., a p or an np chart — should be used.

In such cases the use of narrow-limit gages may be profitable. In this technique gages are prepared with limits stricter than the product specifications. This modification, if applied properly, may strengthen the method against a shift of the process or an increase of the variance.

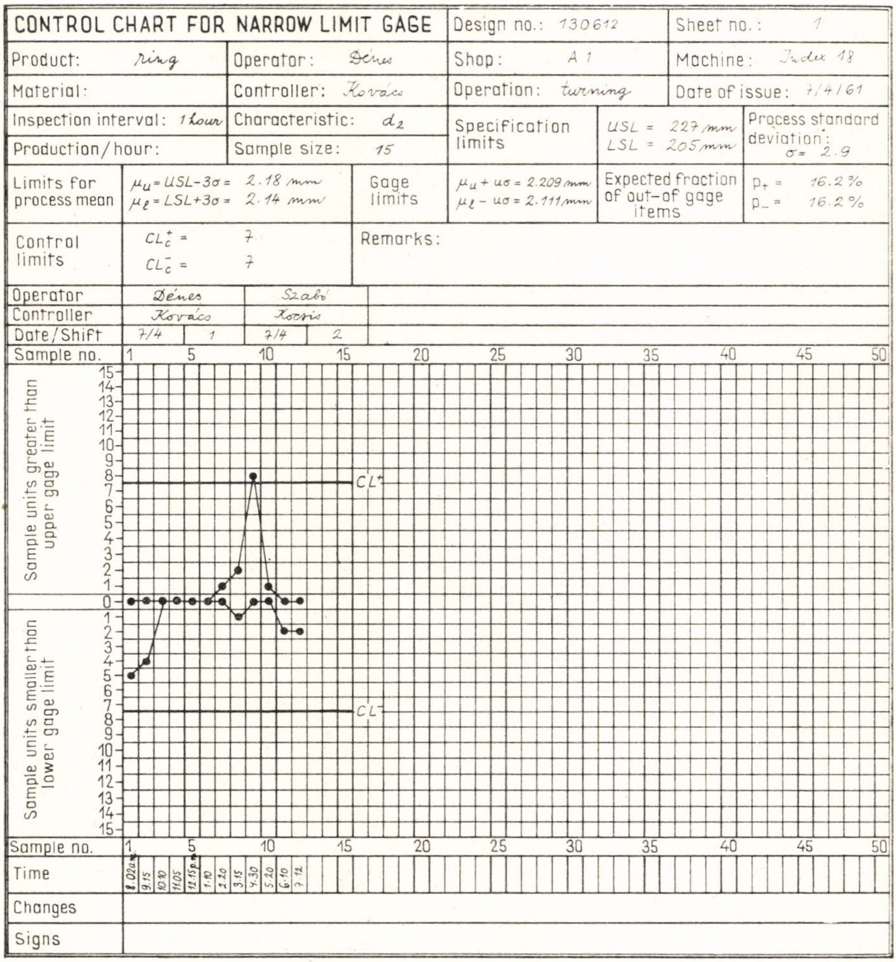
The use of narrow gages eliminates a major disadvantage which is usually common to attribute control methods: they require rather large sample sizes, particularly large samples if the fraction defective is small. In the case of narrow limit gages the part of the fraction defective is taken over by the fraction of out-of-gage items, and the latter can be made big enough by the appropriate choice of the gage limits.

It is suitable to begin the introduction of this method by performing a preliminary analysis in which the method by variables is used. From these data we may decide on the gage setting; subsequently, the calculation of the control limits may be based on the same data.

(A) In the majority of cases where we decide to introduce this method, the tolerance field is substantially larger than the process capability, and hence, the process mean is allowed to move within an interval $[\mu_l, \mu_u]$ (Section 1.2.1.3). In 1.2.3.2 we suggested the use of modified control limits under such circumstances; but narrow-limit gaging may be applied if a simpler method is preferred.

A disadvantage of this method is that it presupposes the stability of the variance and the shape of the underlying distribution, while on the other hand it allows no possibility of controlling them. If the variance increases, the probability of the error of the second kind will increase, whereas a decrease in process variance increases the probability of the error of the first kind.

For the determination of μ_l and μ_u we refer to Section 1.2.3.2. Once they are calculated, the gage limits are to be set at the points $\mu_u + u\,\sigma$ and $\mu_l - u\sigma$. Here σ denotes the process standard deviation. The suitable choice of the value of u will determine the fraction of the out-of-gage items. Assuming the worst admitted case, i.e., that the process mean is at μ_u or at μ_l, we may compute this fraction with the help of the distribution function of the basic variable. If this distribution is normal, for example, the fraction out-of-gage is $\Phi(-u)$ in both cases (in the respective directions). From the value of p, the limit UCL_c can be calculated according to



CONTROL CHART FOR NARROW LIMIT GAGE		Design no.: 130612	Sheet no.: 1
Product: ring	Operator: Dénes	Shop: A 1	Machine: Index 18
Material:	Controller: Kovács	Operation: turning	Date of issue: 7/4/61
Inspection interval: 1 hour	Characteristic: d_2	Specification limits	$USL = 227$ mm $LSL = 205$ mm
Production/hour:	Sample size: 15		Process standard deviation: $\sigma = 2.9$

Limits for process mean	$\mu_u = USL - 3\sigma = 2.18$ mm $\mu_l = LSL + 3\sigma = 2.14$ mm	Gage limits	$\mu_u + u\sigma = 2.209$ mm $\mu_l - u\sigma = 2.111$ mm	Expected fraction of out-of gage items	$p_+ = 16.2\%$ $p_- = 16.2\%$
Control limits	$CL_c^+ = 7$ $CL_c^- = 7$	Remarks:			

Operator	Dénes		Szabó	
Controller	Kovács		Kocsis	
Date/Shift	7/4	1	7/4	2

Sample no.	1	2	3	4	5	6	7	8	9	10	11	12
Time	8.02 a.m.	9.15	10.10	11.05	12.15 p.m.	1.10	2.20	3.15	4.30	5.20	6.10	7.12

Changes

Signs

Sample no.
1
5
10
15
20
25
30
35
40
45
50
Sample units greater than upper gage limit
Sample units smaller than lower gage limit
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
CL^+
CL^-

FIG. 49

the rules described in 1.3.2 (i.e., by using Table 11), for any sample size n, provided that the condition of small p is fulfilled. If $p > 10\%$, then the Poisson approximation is insufficient and leads to broader limits than intended. A table of the binomial distribution, e.g. [C 11, C 13], may be used, or we may use the normal approximation:

$$\mathrm{UCL}_c = np + 3.09\sqrt{np(1-p)}\,. \tag{2}$$

The lower control limit may be omitted but UCL_c should be used to control the numbers of units exceeding the upper or lower limits. For this purpose a double np chart is to be designed (Fig. 49). The numbers of units outside the upper and lower gage limits are to be plotted upwards and downwards from the same origin. The control limit should be marked on both sides. Naturally, if for some reason different u values (u^+ and u^-) have been used at the upper and lower gage settings, the two corresponding control limits (UCL_c^+ and UCL_c^-) must be calculated separately.

Many considerations must be taken into account in the choice of u. Obviously, it is desirable that np be reasonably high at the same time when n is not too high.

TABLE 12

NARROW LIMIT GAGING; NORMAL DISTRIBUTION
Tolerance substantially broader than process capability, or unilateral tolerance

Unilateral probability level: 99.9%

Sample size n	Factor for gage setting u	Expected fraction of out-of gage items, p %	Control limit UCL_c
2	1.87	3.1	1
3	2.10	1.8	1
4	1.52	6.4	2
5	1.66	4.8	2
6	1.32	9.4	3
7	1.42	7.8	3
8	1.17	12	4
9	1.08	14	4
10	1.07	14.1	5
11	1.19	12.5	5
12	0.99	16.2	6
13	1.07	14.5	6
14	0.93	17.5	7
15	0.99	16.2	7
16	0.88	19.0	8
17	0.93	17.8	8
18	0.83	20.2	9
19	0.74	22.8	9
20	0.80	21.2	10

Note. Gage limits are to be set at the points $\mu_u + u\sigma$ and $\mu_l - u\sigma$.

One important result which may be considered is due to Cavé (B 14, and A 6).

According to this result, in the case of a choice between two simple hypotheses with equal probabilities of errors of both kinds, the condition $\mathrm{UCL}_c = (n - 1)/2$ (or $\mathrm{UCL}_c = n/2$) ensures an optimum (for fixed n and error probabilities, the distance of two hypotheses will be minimized). In particular (in equating the "errors" to 50%), this means that the slope of the power function at the ordinate point 1/2 will be maximized.

Our Table 12 is based on the results of Cavé. It gives the necessary values for $n \leqslant 20$; odd sample sizes are preferable. The (one-sided) probability level is 99.9%.

If $n > 20$, the value, of p can be determined with the help of the table of the binomial distribution, or formula (2); hence the value of n can be determined with the help of the table of the normal distribution function (Table IV, p. 385).

It is known that the problem of choosing μ_u (μ_l) is connected with the value of the fraction defective contained in the null hypothesis. This value is termed acceptable quality level (AQL, see acceptance sampling, Section 2.2.1). If $\mu_u = \mathrm{USL} - 3\,\sigma(\mu_l = \mathrm{LSL} + 3\,\sigma)$, this corresponds to the value AQL = 0.13%, under the assumption of normality. Some tables—starting from AQL—give the distance of the narrow limit from the specification limit directly. (See [A 27], pp. 13/65—66; and [A 6], p. 528.) Even if such tables are used, the determination of μ_u and μ_l with the help of the table of the normal distribution function may be informative and useful.

The basic principle of Cavé's solution, i.e., regarding both errors equally, may be objectionable on the basis of practical considerations. In practice, small deviations from the null hypothesis are more frequent than those lying in the vicinity of the alternative hypothesis (if the latter is defined according to the prescribed value of the error of the second kind).

The test is performed repeatedly; therefore it is more important to increase the slope of the power function in the neighborhood of the null hypothesis than far away from it. These arguments lead to lower u values (higher UCL_c values) than those given in Table 12.

On the other hand, the nearer the narrow limits are to the specification limits, the less sensitive is the procedure against departures from normality (or the supposed shape of the distribution) and against changes in the variance. Sometimes gages already at our disposal may influence our choice. If the value of u is chosen without using Table 12, $p = \Phi(-u)$

can be determined from Table IV (p. 385), while formula (1), (2) or the binomial table is to be used for computing UCL_c.

The above principles apply almost without change in cases of one-sided tolerance and gage.

(B) The principles are different if there is a two-sided specification and no allowance is given for drift in the process average: the process is to be kept in control in the strict sense (the capability is not larger than the tolerance). In this case narrow limits are able to control both the average and the variation of the process.

In the remaining parts of this section we shall use the following notations, slightly departing from the foregoing:

n sample size
c^+ number of sample units outside the upper gage limit
c^- number of sample units outside the lower gage limit
$c = c^+ + c^-$
$c^* = c^+ - c^-$
μ process average
σ process standard deviation
u^+ the upper gage limit is $\mu + u^+\sigma$
u^- the lower gage limit is $\mu - u^-\sigma$
$p^+ = \Phi(-u^+)$
$p^- = \Phi(-u^-)$
$p = p^+ + p^-$
UCL_c^+, UCL_c^-, UCL_c, UCL_c^* (upper) control limit for c^+, c^-, c and c^*, respectively
LCL_c^* lower control limit for c^*
$\alpha(= 0.001)$ unilateral significance level used in the chart for c (in the chart for c^* the bilateral significance level 2α is used).

In the design of the control chart, its double purpose is to be taken into account.

In controlling the process average, the bilateral chart is to be used as in case (A). If the sample size is so large that the expectation of the out-of-gage items in one sample becomes rather large (larger than 20–30), then we may plot a control chart for c^* instead. For the computations of the limits we may use the normal approximations, in this case

$$\mathrm{UCL}_c^* = n(p^+ - p^-) + 3.09 \sqrt{n[p - (p^+ - p^-)^2]}$$

$$\mathrm{LCL}_c^* = n(p^+ - p^-) - 3.09 \sqrt{n[p - (p^+ - p^-)^2]}$$

or if $p^{+} = p^{-}$:

$$\mathrm{UCL}_c^{*} = 3.09\sqrt{np}$$

$$\mathrm{LCL}_c^{*} = -\,3.09\sqrt{np}.$$

In order to control the process variance, a separate control diagram for the total number of out-of-gage items is to be plotted on the chart (or the distance of the point on the bilateral chart is to be compared with UCL_c).

Below, we deal with the optimization problems relative to c and c^* charts in the case of the normal distribution.

Each of the solutions of these two problems yields just one consideration for the choice of the sample size. This choice is, as in the former case, a matter of compromise between several requirements.

Unlike the problem treated in (A) (which, in the present notation, relates to the charts for c^{+}, c^{-}), these two problems do not admit easy solutions. Here we give asymptotic solutions; alternatively, they may be precisely solved by numerical methods.

(a) Maximization of the power against variance increase (chart for c). Let μ and $\sigma/(1-\delta)$ be the expectation and standard deviation, respectively, under the alternative hypothesis. The probability that a single measurement will lie between the gage limits is in this case,

$$1 - p_1 = \Phi(u^{+}(1-\delta)) - \Phi(-\,u^{-}(1-\delta)).$$

c may be regarded as normally distributed with expectation np_1 and variance $np_1(1-p_1)$. Therefore, it lies below its control limit $x = \mathrm{UCL}_c$ with probability (with δ sufficiently small and n sufficiently large)

$$\Phi\left(\frac{x-np_1}{\sqrt{np_1(1-p_1)}}\right) \approx \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right) + \delta\cdot\left[\frac{d\Phi\left(\dfrac{x-np_1}{\sqrt{np_1(1-p_1)}}\right)}{d\,\delta}\right]_{\delta=0} \approx$$

$$1-\alpha-\frac{\left[(u^{+}\varphi(u^{+}) + u^{-}\,\varphi(u^{-})\right]\varphi(\Phi^{-1}(\alpha))\,\delta\sqrt{n}}{\sqrt{\left[\Phi(-u^{+}) + \Phi(-\,u^{-})\right]\left[1-\Phi(-\,u^{+}) - \Phi(-\,u^{-})\right]}}$$

i.e.

$$\frac{u^{+}\,\varphi(u^{+}) + u^{-}\,\varphi(u^{-})}{\sqrt{\left[\Phi(-\,u^{+}) + \Phi(-\,u^{-})\right]\left[1-\Phi(-\,u^{+}) - \Phi(-\,u^{-})\right]}}$$

is to be maximized. (Note that $u^+ + u^- > 0$ holds by definition.) If u is defined by the equation $\Phi(-u^+) + \Phi(-u^-) = 2\Phi(-u)$, it can be proved that the inequality

$$u^+\varphi(u^+) + u^-\varphi(u^-) \leqslant 2u\varphi(u)$$

is true, with equality holding only in the case $u^+ = u^- = u$. Hence our task reduces to the maximization of the function

$$\frac{u\varphi(u)}{\sqrt{\Phi(-u)\,[1 - 2\,\Phi(-u)]}}\,.$$

Equation

$$u - \frac{1}{u} = \frac{\varphi(u)}{2\,\Phi(-u)} - \frac{\varphi(u)}{1 - 2\,\Phi(u)}$$

leads to the solution of our problem; we obtain $u = 1.48$, i.e., $p = 15.8\%$. This is an asymptotic solution. In the light of the result given in (A) it seems advisable to obtain an approximation in the finite case by putting

$$\frac{x+1}{n+1} \approx 15.8\ \%$$

and determining p and u from the tables of the binomial and normal distribution functions, respectively. This solution is asymptotically equivalent to the former one, and is the solution to the following modification of the original problem:

Denoting by $Q(\delta)$ the power function of the procedure, what choice of u will maximize $Q'(\delta_0)$ where δ_0 is defined by $Q(\delta_0) = 1/2$? In order to solve this modified problem, we have to use the central limit theorem only in the vicinity of the expectation of the binomial distribution; as is known, the approximation is better here than in the tails.

The solution $\mathrm{UCL}_c \approx 0.158\ (n + 1) - 1$ can be regarded as a first approximation and, for given n, it is not difficult to find the exact optimum by the trial-and-error method, using the tables mentioned. However, as u is chosen somewhat arbitrarily taking into consideration several points of view, the necessity of an exact solution is questionable.

(b) Maximization of the power against shift if a chart for the variable c^* is used. Let the expectation and the standard deviation of the individual measurement be $\mu + \sigma\delta$, σ, respectively, in the alternative hypothesis. Thus, the probability that a single unit exceeds the upper gage limit is equal to

$$p_1^+ = \Phi(-u^+ - \delta),$$

while the same probability regarding the lower gage limit is

$$p_1^- = \Phi(-u^- - \delta).$$

As an approximation c^* can be regarded as to be normally distributed with expectation $n(p_1^+ - p_1^-)$ and standard deviation $\sqrt{n[p_1 - (p_1^+ - p_1^-)^2]}$, where $p_1 = p_1^+ + p_1^-$. Therefore it remains between its control limits $x_u = UCL_0^*$ and $x_l = LCL_c^*$ with probability

$$\Phi\left(\frac{x_u - n(p_1^+ - p_1^-)}{\sqrt{n[p_1 - (p_1^+ - p_1^-)^2]}}\right) - \Phi\left(\frac{x_l - n(p_1^+ - p_1^-)}{\sqrt{n[p_1 - (p_1^+ - p_1^-)^2]}}\right) \approx$$
$$1 - 2\alpha + \frac{2[\varphi(u^+) + \varphi(u^-)]\varphi(\Phi^{-1}(\alpha))\delta\sqrt{n}}{\sqrt{\Phi(-u^+) + \Phi(-u^-) - [\Phi(u^+) - \Phi(-u^-)]^2}},$$

i.e., the expression

$$\frac{\varphi(u^+) + \varphi(u^-)}{\sqrt{\Phi(-u^+) + \Phi(-u^-) - [\Phi(-u^+) - \Phi(-u^-)]^2}}$$

is to be maximized. We may suppose that $u^- > 0$, since otherwise the above expression could be augmented by putting $-u^-$ instead of u^-. If $u > 0$ is defined by the formula

$$\varphi(u^+) + \varphi(u^-) = 2\varphi(u),$$

it can be proved that the inequality

$$\Phi(-u^+) + \Phi(-u^-) - [\Phi(-u^+) - \Phi(-u^-)]^2 \geqslant 2\Phi(-u)$$

holds, with equality occurring only in the case $u^+ = u^- = u$. Hence we may restrict our investigation to maximizing the expression $\varphi(u)/\sqrt{\Phi(-u)}$. We obtain by means of the equation $u = \varphi(u)/[2\Phi(-u)]$ the following solution to our problem: $u = 0.61$, i.e., $p^+ = 0.27$.

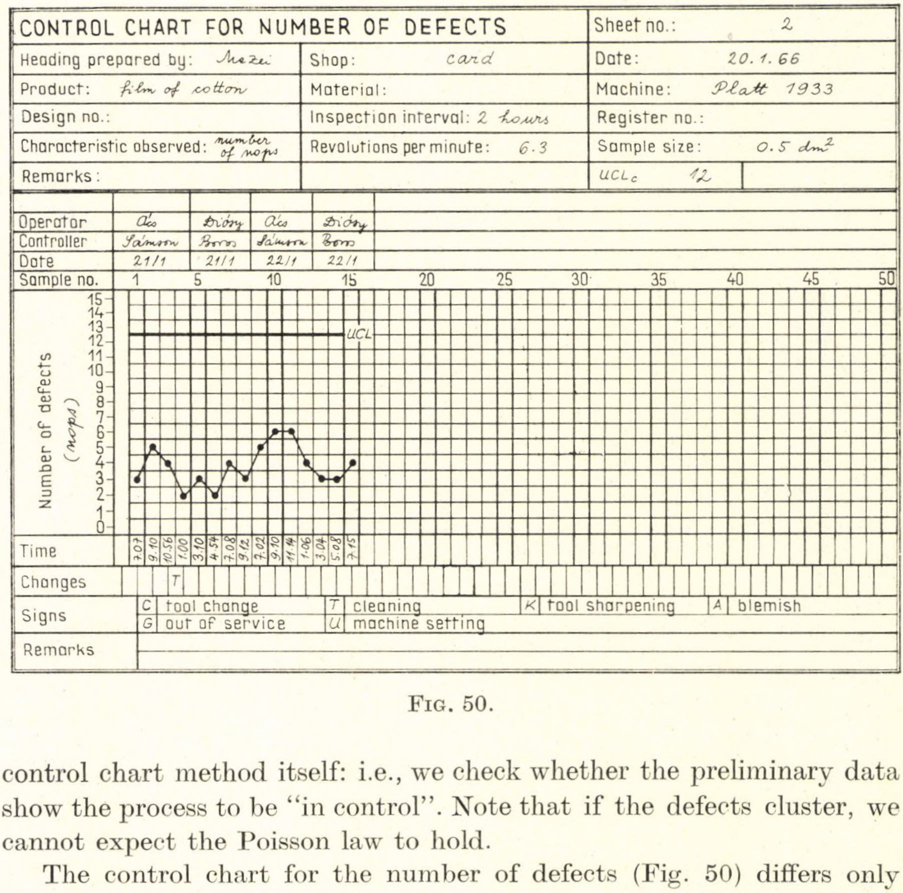
It is probable that, for finite n, the optimal value of u is a decreasing function of n, as in the former cases.

1.3.5 Control Charts for the Number of Defects

Instead of qualifying a sample unit as being good or bad, we may consider the number of defects contained in it. In many cases the latter gives a more appropriate characterization of the quality of the article. Such defects may be, for example: surface blemishes, loose screws, exposed wires.

In such circumstances the *control chart for number of defects* is applied, especially if the distribution of the number of defects can be approximated by the Poisson law. (For other situations see the next section.)

In a large number of cases this condition is practically assured by the relative homogeneity of the production process, in other cases by small sample units taken from the production by a representative method. It is, however, the task of the preliminary analysis to detect significant departures from the Poisson law, if such exist. For this either the Poisson index of dispersion may be used (see, e.g. Haight [A 22], p. 94), or the control chart method itself: i.e., we check whether the preliminary data show the process to be "in control". Note that if the defects cluster, we cannot expect the Poisson law to hold.



CONTROL CHART FOR NUMBER OF DEFECTS		Sheet no.: 2
Heading prepared by: Mezei	Shop: card	Date: 20.1.66
Product: film of cotton	Material:	Machine: Platt 1933
Design no.:	Inspection interval: 2 hours	Register no.:
Characteristic observed: number of nops	Revolutions per minute: 6.3	Sample size: 0.5 dm^2
Remarks:		UCL_c 12

Operator	Ács	Dióry	Ács	Dióry
Controller	Sámson	Boros	Sámson	Boros
Date	21/1	21/1	22/1	22/1

Time: 7.07, 9.10, 10.56, 1.00, 3.10, 4.54, 7.08, 9.12, 7.02, 9.10, 11.14, 1.06, 3.04, 5.08, 7.15

Changes: T

Signs: C tool change; T cleaning; K tool sharpening; A blemish; G out of service; U machine setting

Remarks

FIG. 50.

The control chart for the number of defects (Fig. 50) differs only slightly from the control chart for defectives. Let us mention the following points:

(a) The number of defects may surpass the sample size n.

(b) This control chart is a specific tool for "continuous materials", i.e., materials which do not consist of pieces of equal size. In such cases the choice of a "unit" might be a matter of an arbitrary decision. However, once defined, it is important that the inspection unit be kept constant. In the case of such products, the sample size may take on non-integer values.

(c) For the amount of preliminary data and for the sample size, the same principles are relevant as in the case of defectives. Similarly, Table 11 is to be used for computing the control limits. We look for c, the average number of defects per sample which has been calculated from the preliminary data. If c is not contained in the Table, then we use the formulas

$$UCL_c = c + 3.09 \sqrt{c}$$

$$UCL_c = c - 3.09 \sqrt{c}$$

(d) In the case of variable sample size we have again three alternatives, depending on the magnitude of the fluctuations:

α) If the fluctuations are only very slight, we make no alterations but compute the control limits on the basis of the average sample size.

β) If the fluctuations are higher but still rather moderate, we introduce the *control chart for defects per unit* and compute the control limits on the basis of the average sample size.

γ) If the fluctuations are high, we use the control chart for defects per unit, the control limits separately calculated for each sample.

The control limits for defects per unit are to be calculated in the same way as those for the fraction defective (Section 1.3.3).

1.3.6 Other Control Charts for Attributes

As in the case of control charts by variables (III.1.2.3.4), we close the treatment of control charts for attributes by mentioning briefly some of their special types.

(a) The *two sided control chart for the number of defectives* has been discussed in the section on narrow-limit gaging (III.1.3.4). As a matter of fact, this type of chart can be used generally in cases of gage control, irrespective of whether or not the gage is set at narrow limits.

(b) The *chart for demerits* is a modification of the chart for the number of defects. Often several types of defects may occur with different scales of importance attached to them. Accordingly, they have to be taken into consideration with certain weights.

Let $c_1, c_2, \ldots, c_r$ denote the average number of the different types of defects observed in the preliminary (base) period, where the corresponding weights are $w_1, w_2, \ldots, w_r$, respectively. Assuming that the defects are independent and each of them appears according to the Poisson law, the control limits for the number of demerits (weighted sum of the number of defects) may be calculated by the following formulae:

$$UCL_d = \sum_{i=1}^{r} w_i c_i + 3.09 \sqrt{\sum_{i=1}^{r} w_i^2 c_i}$$

$$LCL_d = \sum_{i=1}^{r} w_i c_i - 3.09 \sqrt{\sum_{i=1}^{r} w_i^2 c_i}.$$

(c) *The cumulative sum chart for attributes* is similar to the corresponding type for variables, dealt with in Section 1.2.3.4. (f). For more details, see Johnson and Leone [A 25], pp. 329–39.

(d) If the distribution cannot be approximated by the Poisson law, the empirical variance (or average range) may be calculated from the preliminary data. For the calculation of control limits, see the control chart for individual measurements (Chapter III, Section 1.2.3.2).

1.4 THE CHOICE OF THE INTERVALS BETWEEN TWO CONSECUTIVE CONTROLS

1.4.1 Some Points of View on the Choice of the Length of the Interval

As already mentioned on several occasions, control during production is carried out in practice usually at intervals of equal length, generally of one-half to two hours. In most cases the control at constant intervals proves to be suitable.

The length of the time interval depends on many factors which are mainly of local character. Perhaps this is the reason why this question was not treated for a long time in the literature. The problem of the appropriate choice of control frequency was considered only recently, first in connection with reliability theory. In this field the replacement of units is the analogous problem.

In many industrial or other organizational problems a decision depends on many circumstances, and therefore it is not easy to find an optimal solution, or even an approximately optimal one. Furthermore, most of the factors (circumstances) are closely related to local conditions, they are of different importance at different places. Consequently, a complete elaboration of such questions is neither very expedient nor even possible.

It is much more advisable to give the method for carrying out such an optimality calculation. For this reason we shall give such a calculation in very detailed form in the next Section. Since our considerations will be based on costs which are generally the most important ones, our example may find a large field of application. This detailed procedure will be carried out for the case of a constant interval. The case of varying intervals, which is more complicated, will be considered from a slightly different point of view and only the main features of the procedure will be described.

Finally, we remark that a third possibility would be to choose intervals of random length, according to some fixed random mechanism. But it can be shown (see [B 43], pp. 97 and 107), that for any random inspection plan there is a better nonrandomized one which is better.*

It is also shown (in [B 43]) that the constant length is optimal if the distribution of a failure in the machine is exponential. The constant interval is used in most cases because it is very practical. An exception is a distribution with a so-called monotone hazard rate, which basically offers the use of a monotonely variable interval.

1.4.2 The Case of Time Intervals of Constant Length

We shall consider a production process supervised by statistical quality control methods at intervals separated by time T, the appropriate value of which is to be determined. When the supervisor detects a failure (i.e., he observes a sample characteristic outside the control limits) he sees to an immediate standstill and repair. After this, the machine works until defectives are discovered (or rather until the defective amount exceeds the tolerated quantity). We assume that if the machine once fails, then it goes on with the production of wrong items until the supervisor detects the fault. In connection with this model, the following quantities will play a role:

α the probability of an error of the first kind, i.e., the probability that a working machine will be found by the supervisor to be working badly;

β the probability of an error of the second kind, i.e., that the supervisor does not detect the failure in the production;

---

* The present treatment of the problem is analogous to that of L. C. Hunter given in the cited place and is due to G. Tusnády.

X the time elapsing from a repair (machine setting) until the next occurrence of the failure, i.e., the time during which the machine is working correctly;

A the cost caused by an error of first kind (the machine will be stopped and adjusted needlessly);

B the average cost of restoring of the machine (repair and setting of the machine, new parts needed, etc.), including the loss in time which is necessary for repair (these time intervals will not be taken into consideration);

C the cost of ordinary inspection;

D the loss per time unit caused by defective work (loss of time, material etc.).

Let us consider that the machine begins to work after being repaired at time T_0. The next failure will occur at time $T_0 + X$, while supervisions will take place at $T_0 + kT$, $k = 1, 2, \ldots$. Let us denote by $\varkappa_1$ the integer for which

$$\varkappa_1 T < X < (\varkappa_1 + 1)T,$$

i.e., $\varkappa_1$ is the random variable with the property that the work of the machine will change from good to defective in the interval $(T_0 + \varkappa_1 T, T_0 + (\varkappa_1 + 1)T)$. Maybe the supervisor does not detect the failure at the next control, but only at the inspection $T_0 + (\varkappa_1 + \varkappa_2)T$. In the meantime the production will be defective during an interval σ, for which

$$X + \sigma = (\varkappa_1 + \varkappa_2)T = \varkappa T = \tau.$$

The random variable τ is now a period in the work of the machine from one restoration to the next (forgetting the time needed for restoration).

Taking into account the above definition, the total cost for a period τ is

$$\chi = A\pi + B + C\,\varkappa + D\sigma,$$

where π denotes the number of errors of the first kind during the period X·

Assuming that the machine carries on work for an indefinite but long period, it is reasonable to minimize the average cost per time unit determined by the function

$$K(T) = \frac{E(\chi)}{E(\tau)}\,.$$

First, we turn to the determination of the numerator and the denominator:

$$E(\chi) = AE(\pi) + B + CE(\varkappa_1 + \varkappa_2) + DE(\sigma);$$

$$E(\tau) = E(\chi) + E(\sigma) = TE(\varkappa_1 + \varkappa_2).$$

The expected value of $\varkappa_1$: Let $F(x)$ be the distribution function of X: $P(X \leqslant x) = F(x)$, $x \geqslant 0$. Then for the distribution of $\varkappa_1$ we have:

$$P(\varkappa_1 = i) = P(iT < X \leqq (i+1)T) = F((i+1)T) - F(iT),$$

$$i = 0, 1, 2, \ldots$$

and by a slight computation we may obtain

$$\sum_{i=0}^{\infty} P(\varkappa_1 = i) = 1$$

$$\begin{aligned} E(\varkappa_1) &= \sum_{i=1}^{\infty} iP(\varkappa_1 = i) = \sum_{i=1}^{\infty} (i+1)P(\varkappa_1 = i) - 1 = \\ &= \sum_{i=1}^{\infty} i[F(iT) - F((i-1)T)] - 1 = \\ &= \sum_{i=0}^{\infty} (1 - F(iT)) - 1 = m(T) - 1, \end{aligned}$$

where the notation $m(T) = \sum_{i=0}^{\infty} (1 - F(iT)) = E(\varkappa_1) + 1$ is used.

The expected value of $\varkappa_2$: the probability of an error of the second kind was β; the event $\{\varkappa_2 = i\}$ means that the controller was not able to discover the failure at $(i - 1)$ subsequent controls, but at the ith occasion he discovered it. Consequently

$$P(\varkappa_2 = i) = \beta^{i-1}(1 - \beta), \quad i = 1, 2, \ldots$$

from which

$$E(\varkappa_2) = \sum_{i=0}^{\infty} i\beta^{i-1}(1 - \beta) = \frac{1}{1 - \beta}.$$

The expected value of π: this depends on the number of supervisions made during interval X; if this number is i, then the expected value of π equals $i\alpha$. Therefore—using the theorem of total probability—

$$E(\pi) = \sum_{i=1}^{\infty} E(\pi)\big|\varkappa_1 = i)P(\varkappa_1 = i) = \sum_{i=1}^{\infty} \alpha iP(\varkappa_1 = i) = \alpha E(\varkappa_1) = \alpha(m(T) - 1)$$

holds.

The expected value of σ: Using the notations

$$E(X) = E, \quad \frac{\beta}{1-\beta} = \varepsilon\,,$$

we obtain: as $\sigma = (\varkappa_1 + \varkappa_2)T - X$,

$$E(\sigma) = T(E(\varkappa_1) + E(\varkappa_2)) - E(X) = T(m(T) + \varepsilon) - E.$$

Now for the expected value of χ and τ we obtain

$$E(\chi) = (m(T) + \varepsilon)\,(DT + C + A\alpha) - ED + B - A\alpha(\varepsilon + 1),$$

$$E(\tau) = T(m(T) + \varepsilon).$$

Finally

$$K(T) = \frac{E(\chi)}{E(\tau)} = D + \frac{C + A\alpha}{T} - \frac{ED + A\alpha(\varepsilon+1) - B}{T(m(T)+\varepsilon)}\,.$$

Let us introduce the following constants:

$$C_1 = C + A\alpha$$

and

$$H = ED + A\alpha(\varepsilon + 1) - B.$$

The function to be minimized is obtained in the form

$$K(T) = D + \frac{G}{T} - \frac{H}{T(m(T)+\varepsilon)}\,.$$

A necessary condition for the minimum is

$$K'(T) = -\frac{G}{T^2} + H\,\frac{m(T) + \varepsilon + Tm'(T)}{T^2(m(T)+\varepsilon)^2} = 0\,.$$

Denoting now the expression $Tm(T) + \varepsilon T - E$ by $\mu(T)$, this condition can be written in the following simple form:

$$G\left(\frac{\mu(T)+E}{T}\right)^2 = H\,\mu'(T)\,. \tag{3}$$

Knowing the distribution function $F(x)$, this can be solved by ad hoc numerical methods; a simple general solution does not exist. Under a special assumption we shall give an approximation method and then turn to the exponential case.

(α) An approximation for the optimal solution of T. Let us suppose that (1) the time T is small compared with the average time E of faultless work of the machine; i.e., $T \ll E$, and that (2) the expected time of occurrence of a failure is near the middle of an interval T:

$$E(\sigma - (\varkappa_2 - 1)T) \approx \frac{T}{2}.$$

From the relation

$$\sigma - (\varkappa_2 - 1)\, T = (\varkappa_1 + 1)\, T - X,$$

we obtain

$$E(\sigma - (\varkappa_2 - 1)T) = Tm(T) - E = \mu(T) - \varepsilon\, T \approx \frac{T}{2}$$

so that

$$\mu(T) - \varepsilon T = \frac{T}{2}\,, \; \mu'(T) = \varepsilon + \frac{1}{2} = \varepsilon'.$$

Relation (3) has now the following form:

$$G \left(\frac{\varepsilon'\, T + E}{T} \right)^2 = H\, \varepsilon'$$

and hence we obtain the value which—under assumptions (1) and (2)—minimizes $K(T)$ as

$$T_{\min} = \frac{E}{\sqrt{\frac{H}{G}\, \varepsilon' - \varepsilon'}}$$

(β) *The exponential case.* Let

$$P(X < x) = 1 - e^{-\lambda x},\; x \geqslant 0,$$

then

$$E(X) = \frac{1}{\lambda},$$

$$m(T) = \sum_{i=0}^{\infty} (1 - F(Ti)) = \sum_{i=1}^{\infty} e^{-i\lambda T} = \frac{1}{1 - e^{-\lambda T}}$$

and

$$\mu(T) = \frac{T}{1 - e^{-\lambda T}} + \varepsilon\, T - \frac{1}{\lambda}\,.$$

Accordingly, relation (3) has the following form:

$$G \left(\frac{1}{1 - e^{-\lambda T}} + \varepsilon \right)^2 = H \left(\frac{1 - e^{-\lambda T} - T\lambda e^{-\lambda T}}{(1 - e^{-\lambda T})^2} + \varepsilon \right).$$

Knowing the numerical values of λ, G, H and ε, this equation can be solved with prescribed accuracy.

Considering again the case $T \ll E = \frac{1}{\lambda}$, i.e., $\lambda T \ll 1$, and neglecting terms of third or higher degree in $T\lambda$, as well as terms in which the small quantity $\varepsilon^2 = \left(\frac{\beta}{1-\beta}\right)^2$ occurs, we obtain for the optimal value of T the following expression:

$$T = \frac{E}{\sqrt{\frac{H}{G}\varepsilon' + \varepsilon}} + \frac{\varepsilon E}{\frac{H}{G}\varepsilon' + \varepsilon}.$$

1.4.3 The Case of Time Intervals of Variable Length

The quantity to be optimized here differs from the one used in the case of constant intervals between two consecutive inspections. Our aim now is to determine a time schedule $\mathfrak{T} = \{ T_0 = 0,\ T_1,\ T_2,\ \ldots T_i,\}$ for the inspections which will guarantee an optimum for the *average total cost for some specified period* of time. This means that, instead of minimizing the average cost per unit time, the total cost has to be kept as small as possible by the appropriate choice of a system $\mathfrak{T}$. This means that starting from a restoration of the machine or production process at time $T_0 = 0$, inspections are foreseen at successive instants $T_1 < T_2 < \ldots < T_i < T_{i+1} < \ldots$ The check at time T_i takes place if no failure was detected at T_{i-1}. We now deviate from the procedure used in the preceding example in that *no error of the second kind will be taken into consideration* ($\beta = 0$). A failure is assumed to be detected with probability one. In most other respects the same assumptions will be used as before.

In general, it is more economical to minimize the average cost per unit time than the average cost itself; it is, however, much easier to operate with the latter if a variable interval is used and one wishes to obtain an optimal system $\mathfrak{T}$.

Turning to our problem, we recapitulate the assumptions: An error of the first kind causes a loss of amount A and this happens with probability α. The time between failure of the system is a random variable X, having distribution function $F(x)$, and expected value E.

The average cost of restoration (of the machine or of the production process) is B; the cost of an inspection is C; while the cost per unit time caused by the defective operation of the machine is denoted by D.

Let us consider a policy for inspection, i.e., a system

$$\mathfrak{T} = \{T_0 = 0,\ T_1, T_2, \ldots T_i, \ldots\}.$$

Starting the machine after a restoration at time $T_0 = 0$, suppose that a first failure occurs in the course of the $\varkappa$th interval; $\varkappa$ is a random variable for which $T_{\varkappa-1} < X < T_\varkappa$ holds. According to our assumptions, the failure will be detected at the instant $T_\varkappa$. This means that the defective work of the machine will be carried on during a time interval σ for which $X + \sigma = T_\varkappa$ holds.

Denoting now by π the number of errors of the first kind made during this interval X, the total cost for a period will be

$$\chi = A\pi + B + C\varkappa + D\sigma,$$

which is of the same form as before. *Now the expected value of χ is to be minimized* by an appropriate choice of the system $\mathfrak{T}$:

$$K(\mathfrak{T}) = E(\chi) = AE(\pi) + B + CE(\varkappa) + DE(\sigma).$$

As we now have consecutive time intervals of variable length, the expected values of the terms take on a slightly different form.

The expected value of $\varkappa$:

$$P(\varkappa = i) = P(T_{i-1} < X < T_i) = F(T_i) - F(T_{i-1}), \quad i = 1, 2, \ldots$$

and

$$E(\varkappa) = \sum_{i=1}^{\infty} iP(\varkappa = i) = \sum_{i=1}^{\infty} i(F(T_i) - F(T_i{}^{-}{}_1)) = \sum_{i=1}^{\infty} (1 - F(T^i)) = m(\mathfrak{T}).$$

The expected value of π: if there were i inspections during the period, then the expected value of π is $i\alpha$; consequently

$$E(\pi) = \sum_{i=1}^{\infty} E(\pi \mid \varkappa = i)P(\varkappa = i) = \sum_{i=1}^{\infty} \alpha iP(\varkappa = i) = \alpha E(\varkappa) = \alpha m(\mathfrak{T}).$$

The expected value of σ: As $\sigma = T_\varkappa - X$, the relation $E(\sigma) = E(T_\varkappa) - - E$ holds. But $P(T_\varkappa = T_i) = P(T_{i-1} < X < T_i)$; therefore $E(T_\varkappa) = = \sum_{i=1}^{\infty} T_i(F(T_i) - F(T_{i-1})) = m_1(\mathfrak{T})$, and using this notation we have

$$E(\sigma) = m_1(\mathfrak{T}) - E.$$

Therefore, we obtain for the expected value of the cost over a whole period the following expression:

$$K(\mathfrak{T}) = A\alpha m(\mathfrak{T}) + B + Cm(\mathfrak{T}) + D(m_1(\mathfrak{T}) - E) = Gm(\mathfrak{T}) + Dm_1(\mathfrak{T}) + H,$$

where the notations

$$G = A\alpha + C, \qquad H = B - ED$$

have been used. Substituting the values of $m(\mathfrak{T})$ and $m_1(\mathfrak{T})$, we obtain after a slight modification

$$K(\mathfrak{T}) = \sum_{i=1}^{\infty} (Gi + DT_i)\,(F(T_i) - F(T_{i-1})) + H.$$

The problem is more complicated than in the case of constant time interval: $K(\mathfrak{T})$ is now a function of infinitely many variables and it is not at all clear whether an optimum system minimizing $K(\mathfrak{T})$ exists or not. We shall follow the methods of Barlow and Proschan [A 2] and Hunter [B 43]. Their most important statements and theorems are given below. Regarding the proofs and a more detailed treatment of the problem we refer to the cited works where further literature is given.

First of all, the following theorem is true: If $F(x)$ is continuous with finite expected value E, then there exists an optimum procedure

$$\mathfrak{T}^* = \{T_0 = 0, T_1^*, T_2^*, \dots T_i^*, \dots\},$$

which minimizes $K(\mathfrak{T})$.

We consider now the special case when the defective operation of the machine starts no later than $t = T$ (with probability one, i.e., if $F(x) = 1$ for $x \geqslant T$). This is the case, e.g., when the knife wears out after a given time. In such cases an inspection is required at the end of the time interval $(0, T)$. But in certain cases no more checking is necessary. This is contained in the following statement:

If $F(x) = 1$ for $X \geqslant T$, and

$$F(x) \leqslant \frac{1}{1 + \dfrac{D}{G}(T - x)}$$

for $0 \leqslant x \leqslant T$, then the optimum policy consists of a single inspection at time T. On the other hand, if for some value of x the inequality

$$F(x) > \frac{1}{1 + \dfrac{D}{G}(T - x)}$$

holds, then the optimum system $\mathfrak{T}^*$ will require one or more inspections before time T.

Turning to the general case in which an optimum system $\mathfrak{T}^*$ exists, a necessary condition for a system to be optimal is given by the relation

$$\frac{\partial K(\mathfrak{T})}{\partial T_i} = 0, \quad i = 1, 2, \dots .$$

This relation can be expressed in the form

$$T_{i+1} - T_i = \frac{F(T_i) - F(T_{i-1})}{f(T_i)} + \frac{G}{D};$$

where $f(x)$ denotes the density function of $F(x)$. If the first term T_1^* belonging to an optimal system is known, then the remaining terms can be determined by iteration. But, unfortunately, this is not the case in general, and hence the determination of an optimum system is not very simple.

A very important class of distributions is the one whose members have Pólya density functions of order 2. These are characterized by the property that $\log f(x)$ is a concave function. These distributions have a monotonely increasing hazard rate

$$\frac{f(x)}{1 - F(x)} \qquad \text{(see Section 3.5, and [A 2]).}$$

Our next consideration holds for a random variable X having Pólya density function of order 2 (see [B 47]); we shall give a method for the determination of an optimal system in this case.

Let us denote by $\mathfrak{T} = \{T_0 = 0, T_1, T_2, \ldots\}$ an inspection policy, and by $\mathfrak{T}^* = \{T_0 = 0, T_1^*, T_2^*, \ldots\}$ an optimal sequence. Using the notation $\Delta_i = T_i - T_{i-1}$, the following holds: choose a T_1 and determine $T_2, T_3, \ldots$, recursively with the aid of the relation

$$T_{i+1} - T_i = \frac{F(T_i) - F(T_{i-1})}{f(T_i)} - \frac{G}{D};$$

then if $T_i > T_i^*$, $\Delta_i > \Delta_{i-1}$ for some positive integer i; while if $T_i < T_i^*$, then $\Delta_i < 0$ for some positive integer i. Hunter suggests the following choice of T_1:

$$G = D \int_0^{T_1} F(x)\, dx\,.$$

He gives an example for $F(x) = \Phi(x;\ 500,\ 100)$ with $G = 10$, $D = 1$, where $T_1 = 422.4$ hours $< T_1^* < T_1' = 422.5$ hours; as it is shown, $\Delta_9 = 27.40 < \Delta_{10} = 30.00$ for the choice $T_1 = 422.4$, while $\Delta_{15}' = -9.0$ for the choice $T_1' = 422.5$. The suitable time for the first control lies between 422.4 and 422.5 hours.

If $F(x)$ is the uniform distribution on the interval $(0, T)$, then it can be shown that an optimal policy terminates after a finite number—say

n—of inspections and

$$T_i^* = i\frac{T}{n} + i(n-i)\,\frac{G}{ZD}, \quad i = 0, 1, 2, \ldots,$$

where

$$\frac{T}{n} + \frac{G}{2\,D}\,(n - 2\,i + 1) > 0 \text{ for all } i$$

and

$$n(n-1) < 2\frac{D}{G}T$$

must hold.

2 ACCEPTANCE SAMPLING

2.1 ON ACCEPTANCE SAMPLING IN GENERAL

Generally speaking, the purpose of acceptance sampling is to make a decision between two possibilities concerning a lot of products, namely either accepting the lot (as being of good quality) or refusing it (as being of poor quality).

Acceptance sampling is performed by means of a *sampling plan* which is essentially a significance test: it specifies rules governing the decision concerning a lot. Usually, the appropriate sampling plan is selected from a set of plans contained in so-called *sampling tables.*

Acceptance sampling procedures are much simpler than control chart methods. Their influence on the production process is, however, only indirect, they are less effective as a tool for assuring good quality than the control chart methods. Nevertheless—being much simpler to carry out and easier to control by the consumer—their application in the acceptance of lots of products is more widespread.

The opportunities for applying these sampling tables are broader than in acceptance sampling in its narrower sense: sampling tables can be used for the inspection of incoming material and parts as well as for process inspection at various points in the manufacturing operations. The inspection of a finished product may be the basis of an actual acceptance procedure between the producer and consumer; the manufacturer may wish to inspect his own product before selling it, or the purchaser may wish to control a product already accepted by him. In this book we consider acceptance sampling in the broader sense, including any use of sampling tables.

Similarly, statistical literature uses the words "producer" and "consumer" in a broader sense than in their common usage. They may refer to persons, companies, departments, etc. that sell, prepare, buy or accept products or other goods.

There are two main types of acceptance sampling: acceptance by attributes and acceptance by variables. Acceptance sampling by attributes is used in the great majority of cases and in this respect no considerable change is expected in the future.

Apart from the general difficulties of the method based on variables (many quality characteristics are observable only as attributes; inspection is less expensive if gages are used instead of measurements; often there are many important quality characteristics on the same product), the following circumstances give a limitation to its use:

(1) Acceptance sampling is much simpler to perform for attributes than for variables.

(2) The existing tables for sampling by variables assume that the original data are normally distributed, and acceptance sampling is much more sensitive to departures from normality than the control chart methods.

Nevertheless, the use of acceptance sampling by variables has the great advantage that—if applicable—it utilizes the surplus of information contained in quantitative data; hence a sampling plan by variables requires smaller sample size than an attribute plan corresponding to the same OC function.

Most of the basic concepts of acceptance sampling plans and tables will be discussed in Section 2.2 in connection with the simplest form of attribute sampling: *single sampling* (by attributes). This differs from *double, multiple and sequential sampling* in that the final decision is based on the result of one sample only (in the latter types of sampling two or more samples may be taken consecutively, depending on the foregoing results).

2.2 SAMPLING BY ATTRIBUTES

2.2.1 Single Sampling Plan

A single sampling plan specifies the *sample size* n and the *acceptance number* c. If out of the n items at most c turn out to be defective, the lot will be accepted; otherwise it will be rejected. We use the term *rejection number* for $c + 1$; in the terminology of hypotheses testing, this is the critical value of the test statistic.

Given the sample size, the probability level of $1 - \alpha$, and a simple null hypothesis H_0: $p = p_0$ (p is the lot fraction defective, p_0 a specified value of it), we can (approximately) determine the acceptance number as an integer c for which $L(p_0, n, c) \approx 1 - \alpha$. (Or if we look at the probability level as a lower bound, we select the smallest c for which $L(p_0, n, c) \geqslant 1 - \alpha$ still holds.)

Here

$$L(p, n, c) = \sum_{i=0}^{c} \binom{n}{i} p^i (1 - p)^{n-i} \tag{4}$$

is the probability that our sampling plan (we use the notation (n, c) plan) will accept the lot. Formula (4) is valid under the assumption that the sample elements are independent (II.2.9.3). This condition is never completely fulfilled since the sampling is always performed without replacement. If, however, the lot is big enough in relation to the sample (at least about ten times as large) then formula (4) meets all practical requirements. Formula (4) is always suitable for the following modification of the model: let us assume that the defectiveness of any item of the lot is a random event having probability p, these events being mutually independent for the lot elements. This modification is almost always reasonable for practical purposes. If the lots come from a production process, p can be interpreted as the fraction defective for the process (Deming [A 9]).

$L(p, n, c)$ as a function of p is called the *operating characteristic function* (*OC function*) of the plan (n, c) (II. 4.3.3). It is always monotonically decreasing.

Let us note that the composite null hypothesis H_0': $p \leqslant p_0$ leads to the same sampling plans as H_0: $p = p_0$.

In the literature of acceptance sampling the value of p_0 (fraction defective according to the simple null hypothesis) is called *acceptable quality level* (AQL). Concerning a definite sampling plan, AQL means the abscissa at which the OC function takes on the value $1 - \alpha$. A different definition is given in MIL STD 105D [D 1, p. 3]: AQL is the maximum fraction defective that, for purposes of sampling inspection, can be satisfactory as a process average.

The probability of the error of the first kind is called *producer's risk* or α *risk*. If fixed, it is usually 5%, but in some schemes (MIL STD 105D, see Section 2.2.3) it varies depending on the sample size.

If a simple alternative hypothesis is specified, it is called *limiting quality* (LQ) or *lot tolerance proportion defective* (LTPD). The corresponding probability of acceptance (probability of the error of the second kind) is called *consumer's risk* or *β risk*. Its value is usually 10%, sometimes also 5%.

Figure 51 shows the OC function of a single sampling plan.

For any sampling plan (n, c), the values of AQL and LQ can be determined by solving the equations $L(p, n, c) = 1 - \alpha$ and $L(p, n, c) = \beta$, respectively, with respect to p.

Inversely, if the values of AQL and LQ (together with the error probabilities α and β) are given, the respective sampling plan can be determined (approximately) by trial and error. For this purpose tables of the binomial distribution [C 13, C 11] or those of the incomplete Beta function [C 9] may be used. In general this problem has—owing to the discreteness of the binomial distribution—no exact solution. From a practical point of view, however, the problem can in general be solved with sufficient accuracy. The calculation is easier if the Poisson approximation can be applied (Peach and Littauer [B 58]), namely by putting

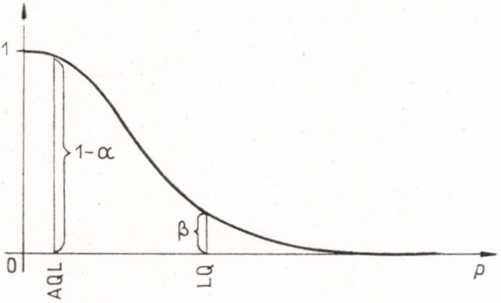


FIG. 51. Operating characteristic function

$$L(p, n, c) \approx \sum_{i=0}^{c} \frac{(np)^i}{i!} e^{-np}.$$

This function does not depend separately on the variables n and p, but only on their product np. If c is given, the equations $L(n, p, c) = 1 - \alpha$ or $L(n, p, c) = \beta$ can easily be solved for np with the help of the table of the Poisson distribution function. The solutions n.AQL and n.LQ are given, for $c \leqslant 10$, in Table 13 ($\alpha = 5\%$, $\beta = 10\%$). The fourth column of the same table shows the ratio of these two quantities, LQ/AQL, which is independent of n.

TABLE 13

DISCRIMINATING RATIO LQ/AQL TO GIVEN ACCEPTANCE NUMBER c

c	n.AQL	n.LQ	LQ/AQL
0	2.30	0.05	0.022
1	3.89	0.36	0.091
2	5.32	0.82	0.154
3	6.68	1.37	0.205
4	7.99	1.97	0.246
5	9.27	2.61	0.282
6	10.53	3.29	0.312
7	11.77	3.98	0.338
8	12.99	4.69	0.361
9	14.21	5.43	0.382
10	15.41	6.17	0.400
11	16.60	6.92	0.417
12	17.78	7.69	0.432
13	18.96	8.46	0.446
14	20.13	9.25	0.459
15	21.3	10.0	0.469
16	22.4	10.8	0.482
17	23.6	11.6	0.491
18	24.7	12.4	0.502
19	25.9	13.2	0.510
20	27.0	14.1	0.522
21	28.2	14.9	0.528
22	29.3	15.7	0.536
23	30.4	16.5	0.543
24	31.6	17.4	0.551
25	32.7	18.2	0.557
26	33.8	19.1	0.565
27	34.9	19.9	0.570
28	36.1	20.7	0.573
29	37.2	21.6	0.581
30	38.3	22.4	0.585

Now, if we wish to construct a sampling plan starting from given values AQL and LQ, we first compute the ratio LQ/AQL and look up the corresponding values of c, n.AQL and n.LQ from the table; hence we obtain n by dividing n.LQ by LQ.

If $p < 10\%$, the Poisson approximation can be considered satisfactory for practical purposes.

2.2.2 Acceptance Sampling Schemes

In the preceding Section we discussed the construction of a sampling plan if the values of AQL and LQ are given in advance.

Usually, sampling inspection tables are based on some principle (strategy) which determines the indexing. This strategy, together with the tables constructed according to it, is called the *sampling scheme*.

Tables of sampling plans indexed by AQL and LQ are given in several places.

Obviously, the indexing of the sampling plans has an influence on the users of the tables when selecting a sampling plan in a given situation. The above-mentioned method of indexing, known as the *two-point scheme*, is one of several possibilities. Experience shows that the use of this scheme leads to uneconomically high sample sizes. This seems to be the reason why this scheme is not widely used in practice. (However, the majority of the sampling inspection tables built up on other principles have auxiliary tables containing the values of these two points).

The idea that economic aspects play a role in the computation of sampling plans may suggest an approach to the problem through *decision theory*. Schemes based on this principle will be mentioned in Section 2.2.8. However, as the necessary preliminary data (prior distribution, costs) are not available in many cases, the application of these tables is considerably limited.

Most widely used are the *AQL-based sampling schemes*, among them the Military Standard 105D (ABC 105D).

As the name shows, the user has to specify the value of AQL. In the preceding Section we have treated the determination of the acceptance number if the sample size and AQL are given. The schemes prescribe the sample size as a function of the lot size, according to a separate principle.

Summarizing, these schemes have the following properties:

(i) The user must specify the AQL and the so-called "inspection level" in advance.

(ii) Within an inspection level, the tables give the sample size as an increasing function of the lot size (with the sampling fraction decreasing, however).

(iii) The tables give the acceptance number corresponding to the sample size, determined according to (ii) and the specified AQL.

More details of this scheme will be given in the next Section.

Some other schemes will be mentioned later. Among them, the Dodge–Romig scheme (Section 2.1.7) is of special importance in cases where refused lots are screened.

2.2.3 Sampling Schemes Based on the Acceptable Quality Level

This section is devoted to sampling schemes based on AQL; we have mentioned these schemes in the preceding Section. Particular attention is given to Military Standard 105D [D 1]. Only simple sampling and the most basic features of the schemes are dealt with in this section—some other properties will be discussed in the forthcoming ones.

Sampling tables of this scheme were first compiled in the United States during the Second World War (See Columbia Sampling Tables, [A 18]).

The most important feature of these schemes is the basic role of the null hypothesis (i.e., of the AQL) in the arrangement of the tables. The predominant role of the AQL as a basis for selecting the most appropriate sampling plan is justified in the majority of the cases, where an acceptance is performed in its real sense (i.e. the vendor and the customer are different organizations).

The reasons for this are the following:

(1) If the average fraction defective (process average) is known from past data, then it is suitable to choose this value as AQL.

The economy of any statistical test depends largely on whether prior information has been taken into account. More precisely, if the prior probability is large enough that the actual value of the fraction defective is near the AQL, the procedure is more economical. It can be viewed then as a substitute for the more accurate Bayesian method. Usually, there is a good chance that the previously experienced quality will be maintained.

This choice of the average fraction defective for AQL is favorable for the vendor since, as long as the formerly reached quality level is maintained, the number of refused lots remains small. (On the other hand the proportion of refused lots increases if the quality deteriorates.)

(2) Although the protection of the customer against the acceptance of a bad lot depends in this case largely on the sample size, the scheme protects the customer in an indirect way as well: it encourages the producer to maintain the quality level at least as high as the AQL. As a matter of fact, if the process quality is poorer than the acceptable quality level, then the proportion of refusals will increase. Thus the

producer has to repair, scrap, or sell his products elsewhere, i.e., he will be involved in increasing expenses. Consequently, he will find it preferable to improve the process quality.

The costs arising from an increasing number of lot refusals and other disadvantages caused by the deterioration of the quality may differ from case to case. Similarly, the costs of improving the quality may differ. Accordingly, the most favorable quality level for the producer depends on his particular circumstances. However, in the majority of cases the sampling plan induces the producer to assure a quality level at least as good as the AQL.

Therefore the AQL should mean for the consumer a level which—as a process average—is acceptable for him. The protection given against worse quality may be increased by the appropriate choice of the inspection level, by the application of the system of tightened inspections (Section 2.2.5), and by the vendor quality rating (Juran [A 27], Chapter 24).

TABLE 14

SAMPLE SIZE CODE LETTERS
(Table I of MIL-STD-105D)

Lot size	Special inspection levels				General inspection levels		
	S-1	S-2	S-3	S-4	I	II	III
2 to 8	A	A	A	A	A	A	B
9 to 15	A	A	A	A	A	B	C
16 to 25	A	A	B	B	B	C	D
26 to 50	A	B	B	C	C	D	E
51 to 90	B	B	C	C	C	E	F
91 to 150	B	B	C	D	D	F	G
151 to 280	B	C	D	E	E	G	H
281 to 500	B	C	D	E	F	H	J
501 to 1,200	C	C	E	F	G	J	K
1 201 to 3,200	C	D	E	G	H	K	L
3 201 to 10,000	C	D	F	G	J	L	M
10 001 to 35,000	C	D	F	H	K	M	N
35 001 to 150,000	D	E	G	J	L	N	P
150 001 to 500,000	D	E	G	J	M	P	Q
500 001 & Over	D	E	H	K	N	Q	R

TABLE 15

(Table II-A of MIL-STD-105D)

SINGLE SAMPLING PLANS FOR NORMAL INSPECTION

Sample size code letter	Sample size	Acceptable quality level (%)												
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A		↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
B		↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
C	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1
D	8	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑
E	13	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓
F	20	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2
G	32	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3
H	50	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4
J	80	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6
K	125	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8
L	200	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11
M	315	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15
N	500	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22
P	800	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑
Q	1 250	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑
R	2 000	↑	↑	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑

Sample size code letter	Sample size	Acceptable quality level (%)												
		4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	0 1		↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	3	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45
C	5	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑
D	8	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	
E	13	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑		
F	20	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑			
G	32	3 4	5 6	7 8	10 11	14 15	21 22	↑						
H	50	5 6	7 8	10 11	14 15	21 22	↑							
J	80	7 8	10 11	14 15	21 22	↑								
K	125	10 11	14 15	21 22	↑									
L	200	14 15	21 22	↑										
M	315	21 22	↑											
N		↑												
P														
Q														
R														

↓ = Use first sampling plan below arrow. If sample size equals, or exceeds lot size, do 100 per cent inspection.

↑ = Use first sampling plan above arrow.

Ac = Acceptance number.

Re = Rejection number.

In the great majority of cases the incentive effect of the acceptance sampling outlined above is of primary importance.

Table 14 (Table I of MIL-STD-105D) gives the code letters which determine the sample size uniquely (see also the second column of Table 15). Table 14 contains seven inspection levels (three "general" and four "special" ones).

In principle, it is desirable that the choice of inspection level be based on cost consideration. It is, however, only in rather exceptional cases that we possess the necessary data to perform this analysis. Information regarding the customary amount of inspection on the same or on similar material, and the experiences obtained from their results may be taken into account. In many cases there are no grounds for any special level. Then the choice of inspection level II is recommended. Hence, this is the most frequently used level. Level III leads to higher and level I to lower sample sizes; the sample sizes given by the special inspection levels are even lower.

Having once chosen the inspection level, we may use the lot size to obtain the code letter from Table 14.

If the code letter (or for some reason, independently of Table 14, the sample size) is determined and the AQL is chosen, we may go to Table 15 (Table II-A of MIL-STD-105D) and obtain the acceptance number. Now the sampling plan is uniquely determined.

(In the table referred to above, the AQL values are given in percentages.)

If we find an arrow instead of an acceptance number in the corresponding cell of the table, then the sample size must be altered. The direction indicated by the arrow should be followed until a non-empty cell is found. The acceptance number contained here should be used together with the sample size given in the same row.

This way of selecting the sampling plan from the table is characteristic of the schemes based on AQL.

As mentioned before, for low percentage values the binomial distribution and the Poisson distribution practically agree. Hence, the plans chosen from the table can be used in both cases i.e. when the number of defective items and when the number of defects are counted, provided that the AQL is not greater than 10%. Those parts of the table where AQL exceeds 10% can be used only in the latter case.

The first schemes based on the AQL tried to maintain a constant probability of 5% for the error of the first kind. In the more recent versions this probability decreases with increasing sample size (cf.

Section 2.1.1); it ranges approximately between 0.10 and 0.003. This change is reasonable because the consumer's risk decreases, anyhow, much more rapidly with sample size in the scheme (if this risk is calculated for a fixed alternative hypothesis).

The Standard also contains special tables for tightened and reduced inspection: these are dealt with in Section 2.2.5.

The Standard gives the LQ values corresponding to the acceptance probabilities 0.10 and 0.05. Values are calculated both on the basis of binomial and Poisson probabilities. For single sampling plans, some values of the OC functions are also tabulated.

These tables facilitate the selection of an appropriate sampling plan according to some principle other than that of the AQL.

Connection between lot size and sample size. With the schemes based on AQL, the sample size in the code tables is dependent on the lot size and on the inspection level. The relationship between the lot size and sample size is nearly the same for the different schemes and is mainly based on heuristic considerations: no strict mathematical justification is given for it (see [A 18], pp. 173—175).

The difficulties of elaborating a relationship based on strict mathematical theory are the same as that of constructing an optimal scheme, (cf. Section 2.2.2). No doubt, n must be an increasing function of the lot size N: if N is large then the loss caused by an erroneous decision is high and the value of the lot is high as well; therefore the costs connected with a larger sample can be afforded.

Sampling plans designed on economic or semi-economic theory lead to relationships which are more or less similar to the one mentioned above (see Sections 2.2.7 and 2.2.8).

2.2.4 Double and Multiple Sampling

2.2.4.1 Double Sampling. In single sampling the number of inspected items is fixed in advance.

In most circumstances, the cost of sampling can be decreased by using double, multiple or sequential sampling instead of a single sample.

In double sampling a first sample of fixed size is always taken and, according to the number of the defectives found in this first sample, one of the following three decisions is made: acception, rejection or taking a second sample.

The quantities which determine a double sampling plan are therefore:

TABLE 16

(Table III-A of MIL-STD-105D)

DOUBLE SAMPLING PLANS FOR NORMAL INSPECTION

Sample size code letter	Sample	Sample size	Cumulative sample size	Acceptable quality level (%)												
				0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5
				Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
B				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
C				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	*
D				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	*	↑
E				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	*	↑	↓
F	First	13	13	↓	↓	↓	↓	↓	↓	↓	↓	↓	*	↑	↓	0 2
	Second	13	26													1 2
G	First	20	20	↓	↓	↓	↓	↓	↓	↓	↓	*	↑	↓	0 2	0 3
	Second	20	40												1 2	3 4
H	First	32	32	↓	↓	↓	↓	↓	↓	↓	*	↑	↓	0 2	0 3	1 4
	Second	32	64											1 2	3 4	4 5
J	First	50	50	↓	↓	↓	↓	↓	↓	*	↑	↓	0 2	0 3	1 4	2 5
	Second	50	100										1 2	3 4	4 5	6 7
K	First	80	80	↓	↓	↓	↓	↓	*	↑	↓	0 2	0 3	1 4	2 5	3 7
	Second	80	160									1 2	3 4	4 5	6 7	8 9
L	First	125	125	↓	↓	↓	↓	*	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9
	Second	125	250								1 2	3 4	4 5	6 7	8 9	12 13
M	First	200	200	↓	↓	↓	*	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11
	Second	200	400							1 2	3 4	4 5	6 7	8 9	12 13	18 19
N	First	315	315	↓	↓	*	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16
	Second	315	630						1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27
P	First	500	500	↓	*	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑
	Second	500	1000					1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	
Q	First	800	800	*	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑
	Second	800	1600				1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27		
R	First	1250	1250	↑	↑	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑
	Second	1250	2500			1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27			

Sample size code letter	Sample	Sample size	Cumulative sample size	Acceptable quality level (%)																									
				4.0		6.5		10		15		25		40		65		100		150		250		400		650		1000	
				Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re
A				↓		*		↓		↓		*		*		*		*		*		*		*		*		*	
B	First	2	2	*		↑				0	2	0	3	1	4	2	5	3	7	5	9	7	11	11	16	17	22	25	31
	Second	2	4							1	2	3	4	4	5	6	7	8	9	12	13	18	19	26	27	37	38	56	57
C	First	3	3	↑		↓		0	2	0	3	1	4	2	5	3	7	5	9	7	11	11	16	17	22	25	31	↑	
	Second	3	6					1	2	3	4	4	5	6	7	8	9	12	13	18	19	26	27	37	38	56	57		
D	First	5	5	↓		0	2	0	3	1	4	2	5	3	7	5	9	7	11	11	16	17	22	25	31	↑			
	Second	5	10			1	2	3	4	4	5	6	7	8	9	12	13	18	19	26	27	37	38	56	57				
E	First	8	8	0	2	0	3	1	4	2	5	3	7	5	9	7	11	11	16	17	22	25	31	↑					
	Second	8	16	1	2	3	4	4	5	6	7	8	9	12	13	18	19	26	27	37	38	56	57						
F	First	13	13	0	3	1	4	2	5	3	7	5	9	7	11	11	16	↑		↑		↑							
	Second	13	26	3	4	4	5	6	7	8	9	12	13	18	19	26	27												
G	First	20	20	1	4	2	5	3	7	5	9	7	11	11	16	↑													
	Second	20	40	4	5	6	7	8	9	12	13	18	19	26	27														
H	First	32	32	2	5	3	7	5	9	7	11	11	16	↑															
	Second	32	64	6	7	8	9	12	13	18	19	26	27																
J	First	50	50	3	7	5	9	7	11	11	16	↑																	
	Second	50	100	8	9	12	13	18	19	26	27																		
K	First	80	80	5	9	7	11	11	16	↑																			
	Second	80	160	12	13	18	19	26	27																				
L	First	125	125	7	11	11	16	↑																					
	Second	125	250	18	19	26	27																						
M	First	200	200	11	16	↑																							
	Second	200	400	26	27																								
N				↑																									
P																													
Q																													
R																													

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100 per cent inspection.
↑ = Use first sampling plan above arrow.
Ac = Acceptance number.
Re = Rejection number.
* = Use corresponding single sampling plan (or alternatively, use double sampling plan below, where available).

TABLE 17

(From Table IV-A of MIL-STD-105D)

MULTIPLE SAMPLING PLANS FOR NORMAL INSPECTION

Sample size code letter	Sample	Sample size	Cumulative sample size	Acceptable quality level (%)												
				0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40
				Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
G	First	8	8	↓	↓	*	**	↓	,, 2	,, 2	,, 3	,, 4	0 4	0 5	1 7	2 9
	Second	8	16						,, 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14
	Third	8	24						0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19
	Fourth	8	32						0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25
	Fifth	8	40						1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29
	Sixth	8	48						1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33
	Seventh	8	56						2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38
H	First	13	13		*	↑	↓	,, 2	,, 2	,, 3	,, 4	0 4	0 5	1 7	2 9	↑
	Second	13	26					,, 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14	
	Third	13	39					0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19	
	Fourth	13	52					0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25	
	Fifth	13	65					1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29	
	Sixth	13	78					1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33	
	Seventh	13	91					2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38	
J	First	20	20	*	↑	↓	,, 2	,, 2	,, 3	,, 4	0 4	0 5	1 7	2 9	↑	
	Second	20	40				,, 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14		
	Third	20	60				0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19		
	Fourth	20	80				0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25		
	Fifth	20	100				1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29		
	Sixth	20	120				1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33		
	Seventh	20	140				2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38		

K	First	32	32	↑		,, 2	,, 2	,, 3	,, 4	0 4	0 5	1 7	2 9	↑	↑	↑
	Second	32	64			,, 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14			
	Third	32	96			0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19			
	Fourth	32	128			0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25			
	Fifth	32	160			1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29			
	Sixth	32	192			1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33			
	Seventh	32	224		↓	2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38			
L	First	50	50		,, 2	,, 2	,, 3	,, 4	0 4	0 5	1 7	2 9	↑			
	Second	50	100		,, 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14				
	Third	50	150		0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19				
	Fourth	50	200		0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25				
	Fifth	50	250		1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29				
	Sixth	50	300		1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33				
	Seventh	50	350	↓	2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38				
M	First	80	80	,, 2	,, 2	,, 3	,, 4	0 4	0 5	1 7	2 9	↑				
	Second	80	160	,, 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14					
	Third	80	240	0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19					
	Fourth	80	320	0 3	1 4	2 5	3 7	5 10	8 13	12 17	25 20					
	Fifth	80	400	1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29					
	Sixth	80	480	1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33					
	Seventh	80	560	2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38					

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100 per cent inspection.

↑ = Use first sampling plan above arrow.

Ac = Acceptance number.

Re = Rejection number.

* = Use corresponding single sampling plan (or alternatively, use multiple plan below).

** = Use single sampling plan for code letter F (or alternatively, use multiple plan below).

,, = Acceptance not permitted at this sample size.

the size of the first sample (n_1); the size of the second sample (n_2); the acceptance number (c_1) and the rejection number (r_1) for the first sample; the acceptance number (c_2) and the rejection number (r_2) for the second sample ($r_2 = c_2 + 1$). These quantities are given in the sampling tables.

The double sampling is performed according to the following steps:

Take a sample of n_1 items. Let d_1 be the number of defectives found in it. Accept the lot if $d_1 \leqslant c_1$; reject the lot if $d_1 \geqslant r_1$; take a second sample of n_2 items if $c_1 < d_1 < c_2$. In the latter case let d_2 be the number of defectives found in the second sample. Accept the lot, if $d_1 + d_2 \leqslant c_2$ and reject it if $d_1 + d_2 \geqslant r_2 = c_2 + 1$.

The operating characteristic function of the double sampling plan is given by the formula

$$L(p) = \sum_{i=0}^{c_1} \binom{n_1}{i} p^i (1-p)^{n_1-i} + \sum_{i=c_1+1}^{r_1-1} \binom{n_1}{i} p^i (1-p)^{n_1-1} \sum_{j=1}^{c_2-i} \binom{n}{j} p^j (1-p)^{n_2-j}.$$

Table 16 (Table III-A of the Military Standard 105D) contains double sampling plans. Double and single sampling plans corresponding to identical code letter and AQL have—approximately—identical OC curves. Accordingly, the probability of accepting a lot is on the whole independent of whether single or double sampling has been performed. The expectation of the inspected sample size is smaller when using the double sampling plan. This means that in some cases double sampling reduces costs (e.g., if inspection cost increases proportionally with sample size).

Also, double sampling has some psychological advantages (borderline lots are given a "second chance" to be accepted).

A disadvantage of double sampling is that its execution is more complicated than that of single sampling.

In Military Standard 105D the first and second samples are the same size. In earlier editions the second sample was twice as large as the first one; this feature turned out to be uneconomical (see Hamaker and van Strik [B 39]). The first (and the second) sample size agrees for each code letter with the sample size of the corresponding single sample belonging to the preceding code letter; thus, e.g., for code letter *H* the sample sizes in double sampling are $n_1 = 32$, $n_2 = 32$, the same as the single sample size for code letter *G*.

2.2.4.2 Multiple Sampling. The extension of the concept of double sampling leads to a type of sampling in which the number of consecutive samples may be more than two. In Military Standard 105D, Table IV-A gives multiple sampling plans. A part of this Table is reproduced in our Table 17. The OC functions are again practically identical to that of the single samping belonging to the same code letter and AQL.

The expectation of the inspected items is always less than in single sampling and almost always less than in double sampling. Consequently, in some cases multiple sampling may be less expensive.

It has, however, the considerable disadvantage of being much more complicated than single and double sampling.

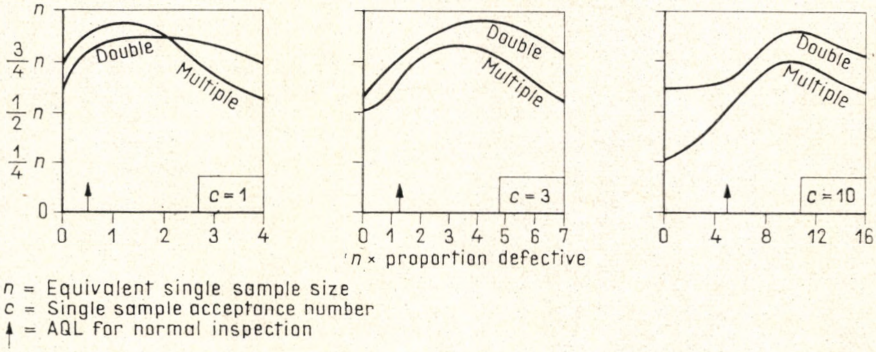


FIG. 52. Average sample size curves (from Table IX of MIL-STD-105D)

In principle, the OC functions of the multiple plans can be obtained in the same way as those of the double plans. Their computation is, however, extremely laborious. The plans contained in the Table just cited can be closely approximated by sequential plans (see Section 2.4).

The curves for the expected sample size in the cases of double and multiple sampling are given in Table IX of the MIL-STD 105D (part of which is reproduced in Fig. 52). The abscissas give the products of the equivalent single sample size and the proportion defective.

Military Standard 105D contains multiple plans of seven equal samples. The sample sizes are equal to the corresponding single sample size belonging to the code letter three steps before. Thus, e.g., the sample size for code letter *P* agrees with the single sample size for code letter *L* (200).

2.2.5 Tightened and Reduced Inspection

An essential feature of MIL-STD-105D (and its predecessors) is its specification of *tightened inspection*, designed to give additional protection to the consumer. As mentioned before, normal inspection (dealt with in

TABLE 18

(Table I-B of MIL-STD-105D)

SINGLE SAMPLING PLANS FOR TIGHTENED INSPECTION

Sample size code letter	Sample size	Acceptable quality level (%)																									
		0.010		0.015		0.025		0.040		0.065		0.10		0.15		0.25		0.40		0.65		1.0		1.5		2.5	
		Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re
A																											
B																											
C																										↓	
D	8																							↓		0	1
E	13																					↓		0	1		
F	20																			↓		0	1			↓	
G	32																	↓		0	1			↓		1	2
H	50															↓		0	1			↓		1	2	2	3
J	80													↓		0	1			↓		1	2	2	3	3	4
K	125											↓		0	1			↓		1	2	2	3	3	4	5	6
L	200									↓		0	1			↓		1	2	2	3	3	4	5	6	8	9
M	315							↓		0	1			↓		1	2	2	3	3	4	5	6	8	9	12	13
N	500					↓		0	1			↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19
P	800			↓		0	1			↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	↑	
Q	1250	↓		0	1			↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	↑			
R	2000	0	1	↑		↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	↑					
S	3150					1	2																				

Sample size code letter	Sample size	Acceptable quality level (%)																									
		4.0		6.5		10		15		25		40		65		100		150		250		400		650		1000	
		Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re
A	2	↓		↓		↓		↓		↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28
B	3	↓		0	1	↓		↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42
C	5	0	1	↓		↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42	↑	
D	8	↓		↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42	↑		↑	
E	13	↓		1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42	↑		↑		↑	
F	20	1	2	2	3	3	4	5	6	8	9	12	13	18	19	↑		↑		↑		↑		↑		↑	
G	32	2	3	3	4	5	6	8	9	12	13	18	19	↑		↑		↑		↑		↑		↑		↑	
H	50	3	4	5	6	8	9	12	13	18	19	↑		↑		↑		↑		↑		↑		↑		↑	
J	80	5	6	8	9	12	13	18	19	↑		↑		↑		↑		↑		↑		↑		↑		↑	
K	125	8	9	12	13	18	19	↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
L	200	12	13	18	19	↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
M	315	18	19			↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
N		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
P		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
Q		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
R		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑		↑	
S																											

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100 per cent inspection.

↑ = Use first sampling plan above arrow.

Ac = Acceptance number.

Re = Rejection number.

TABLE
(TABLE VIII of
LIMIT NUMBERS FOR

Number of sample units from last 10 lots	Acceptable quality level (%)											
	0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5
20 – 29	*	*	*	*	*	*	*	*	*	*	*	*
30 – 49	*	*	*	*	*	*	*	*	*	*	*	*
50 – 79	*	*	*	*	*	*	*	*	*	*	*	*
80 – 129	*	*	*	*	*	*	*	*	*	*	*	*
130 – 199	*	*	*	*	*	*	*	*	*	*	*	0
200 – 319	*	*	*	*	*	*	*	*	*	*	0	0
320 – 499	*	*	*	*	*	*	*	*	*	0	0	1
500 – 799	*	*	*	*	*	*	*	*	0	0	2	3
800 – 1249	*	*	*	*	*	*	*	0	0	2	4	7
1250 – 1999	*	*	*	*	*	*	0	0	2	4	7	13
2000 – 4149	*	*	*	*	*	0	0	2	4	8	14	22
4150 – 4999	*	*	*	*	0	0	1	4	8	14	24	38
5000 – 7999	*	*	*	0	0	2	3	7	14	25	40	63
8000 – 12499	*	*	0	0	2	4	7	14	24	42	68	105
12500 – 19999	*	0	0	2	4	7	13	24	40	69	110	169
20000 – 31499	0	0	2	4	8	14	22	40	68	115	181	
31500 – 49999	0	1	4	8	14	24	38	67	111	186		
50000 & Over	2	3	7	14	25	40	63	110	181	301		

* Denotes that the number of sample units from the last ten lots is not sufficient for reduced inspection used are the most recent ones in sequence, that they have all been on normal inspection, and that none has

the preceding Sections) encourages the producer to keep his process average at least as good as the AQL. It should be remarked, however, that if—in spite of this influence—the quality is moderately poor, the protection against accepting such lots will be unsatisfactory. For this reason additional criteria are needed to protect the consumer if quality history raises suspicion about the quality. In tightened inspection the probability of accepting lots of insufficient quality is decreased; this penalizes the producer for submitting substandard quality and therefore gives him added incentive to improve the quality and bringing it up to the level of AQL.

In Military Standard 105D, Tables IB (reproduced in our Table 18), IIB and IIIB serve for tightened inspection. Shifts from normal to tightened inspection are suggested whenever two out of five consecutive lots have been rejected on the basis of normal inspection (resubmitted lots are to be ignored). If in tightened inspection five consecutive lots have been accepted, then normal inspection is to be reinstituted.

19

MIL-STD-105D)

REDUCED INSPECTION

Acceptable quality level (%)													
2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
*	*	*	0	0	2	4	8	14	22	40	68	115	181
*	*	0	0	1	3	7	13	22	36	63	105	178	277
*	0	0	2	3	7	14	25	40	63	110	181	301	
0	0	2	4	7	14	24	42	68	105	181	297		
0	2	4	7	13	25	42	72	115	177	301	490		
2	4	8	14	22	40	68	115	181	277	471			
4	8	14	24	39	68	113	189						
7	14	25	40	63	110	181							
14	24	42	68	105	181								
24	40	69	110	169									
40	68	115	181										
67	111	186											
110	181												
181													

for this AQL. In this instance more than ten lots may be used for the calculation, provided that the lots been rejected while on original inspection.

According to the instructions of MIL-STD-105D, inspection by these plans has to be stopped if 10 consecutive lots remain in tightened inspection (the number 10 may be altered by the authority).

In this case inspection may switch to LQ-based plans, special efforts may improve the vendor's quality level, the specifications may be revised, or the business relations may even be broken off by the consumer.

The switching rules of the MIL-STD-105D are much simpler than those given in its predecessors. The former rules were based on the average fraction defective. Monte-Carlo investigations have shown that the effectiveness of the new method is practically the same as that of the former [B 55].

Although the tightened inspection with its shifting rules is an effective tool in extending the protection granted for the consumer, it is advisable, in addition, to keep records on the successive sample results and to evaluate them from time to time. The average fraction defective gives

TABLE 20

(Table I-C of MIL-STD-105 D)

SINGLE SAMPLING PLANS FOR REDUCED INSPECTION

Sample size code letter	Sample size	Acceptable quality level (%)																									
		0.010		0.015		0.025		0.040		0.065		0.10		0.15		0.25		0.40		0.65		1.0		1.5		2.5	
		Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re	Ac	Re
A																											
B																										↓	
C	2																							↓		0	1
D	3																					↓		0	1	↑	
E	5																			↓		0	1	↑		↓	
F	8																	↓		0	1	↑		↓		0	2
G	13															↓		0	1	↑		↓		0	2	1	3
H	20													↓		0	1	↑		↓		0	2	1	3	1	4
J	32											↓		0	1	↑		↓		0	2	1	3	1	4	2	5
K	50									↓		0	1	↑		↓		0	2	1	3	1	4	2	5	3	6
L	80							↓		0	1	↑		↓		0	2	1	3	1	4	2	5	3	6	5	8
M	125					↓		0	1	↑		↓		0	2	1	3	1	4	2	5	3	6	5	8	7	10
N	200			↓		0	1	↑		↓		0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13
P	315	↓		0	1	↑		↓		0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13	↑	
Q	500	0	1	↑		↓		0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13	↑			
R	800	↑				0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13	↑					

Sample size code letter	Sample size	Acceptable quality level (%)												
		4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	2	0 1	↑		0 2	1 3	2 4	3 5	5 6	7 8	10 11	14 15	21 22	30 31
C	2	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑
D	3	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑	
E	5	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑		
F	8	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑	↑	↑			
G	13	1 4	2 5	3 6	5 8	7 10	10 13	↑						
H	20	2 5	3 6	5 8	7 10	10 13	↑							
J	32	3 6	5 8	7 10	10 13									
K	50	5 8	7 10	10 13	↑	↑								
L	80	7 10	10 13	↑										
M	125	10 13	↑											
N		↑												
P														
Q														
R														

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100 per cent inspection.

↑ = Use first sampling plan above arrow.

Ac = Acceptance number.

Re = Rejection number.
If the acceptance number has been exceeded, but the rejection number has not been reached, accept the lot, but reinstate normal inspection.

more information about the quality than the proportion of the lots refused. We need this average if we wish to consider the application of reduced inspection (see below). The records on it serve as a good basis for vendor quality rating. They may help the consumer in future decisions concerning business relations, setting of AQL and specifications. If necessary, they may bring—and must bring—about decisions to eliminate chronically poor vendors.

It is advisable to plot the records in a control chart.

Reduced inspection is included in MIL-STD-105D as an adjunct of the sampling scheme (applicable if it "is considered desirable by the responsible authority"). The conditions for shifting to a reduced inspection are the following: the production is at a steady rate; the preceding ten lots have been inspected according to normal inspection and none of them has been rejected; the average proportion defective is significantly lower than the AQL. The last condition is to be controlled by Table 19 (Table VIII of MIL-STD-105D). In some cases this Table prescribes making the decision on more than 10 lots.

If double or multiple sampling is in use, all samples are to be used according to MIL-STD-105D, in estimating the fraction defective. The consideration of the total samples (in contrast to the "first" samples as had been designated in the Columbia Tables) introduces some bias but the estimation is more informative; in addition, it is cautious, since in practical circumstances bias appears in the direction of poorer quality.

In MIL-STD-105D reduced inspection is 40 per cent lower than normal inspection. The plans are given in Tables I-C (reproduced in our Table 20), II-C, III-C of the Standard. In reduced inspection the difference between the acceptance and rejection numbers is larger than 1. If the observed number of defectives falls between these values the lot is accepted, but normal inspection is to be reinstated with the next lot.

The system of normal, tightened and reduced inspections has been analysed by Hald and Thyregod [B 38] and Stephens and Larson [B 76].

2.2.6 Some Further Remarks Concerning Acceptance Sampling

(a) *Formation of inspection lots.* From economical points of view it is desirable that the inspection lots be as large as possible. Larger lots require larger samples, hence they result in better discrimination. On the other hand, the inspection cost per unit is less because the samples contain a smaller ratio of the lot.

We wish to remark, however, that from other aspects smaller inspection lots may be preferable. One of the most important among these aspects is homogeneity: it is undesirable to mix lots of different quality. Many practical points of view (rate of production, availability of storage space) may also influence the decision concerning the formation of lots.

(b) *How to select the sample.* In principle, it is desirable to select the sample by means of random numbers; this is the most reliable way of ensuring random sampling and hence of eliminating subjective biases. This method is, however, seldom practicable.

If random numbers cannot be used, the units of the sample are to be drawn "blind", i.e., without regard to their quality. In the case of large samples proportionate (stratified) sampling is preferable: the lot is divided into sublots and the sample units are selected proportionate to the size of each sublot (within sublots random sampling is performed).

(c) *Procedure in the case of refused lots.* If the lot has been rejected once according to the rules of acceptance sampling, then no resubmission to acceptance procedure is allowed unless all units have been examined and all defective units removed or defects corrected. It is highly desirable to be able to identify lots once rejected. The consumer may consider the use of tightened inspection for the supervision of the rectification. Otherwise, the consumer has insufficient protection against bad quality, since repeated submissions of a bad lot will lead sooner or later to acceptance. If, namely, the value of the OC function equals $L(p)$, where p denotes the true (now unsatisfactory) value of the fraction defective, then out of s independent repetitions of the acceptance sampling at least one will lead to acceptance with probability $1 - (1 - L(p))^s$. For sufficiently large s this will take on values near to 1.

2.2.7 Schemes of Dodge and Romig

The schemes to be discussed in this Section were the first acceptance sampling schemes elaborated. They were developed by H. F. Dodge and H. G. Romig in the twenties and thirties of this century.

Doge and Romig's tables contain two schemes: one based on LQ (called LTPD), and the other on AOQL (see below). Both schemes were worked out principally for the use of manufacturers inspecting their own products. In principle, the user of these schemes makes use of the result of the sample in deciding between the following alternatives: to accept the lot on the basis of the sample, or to carry out a piece-by-piece examination ("screening") with rectification, i.e., the bad items found will be replaced with good ones.

The AOQL scheme aims at guaranteeing the quality of the lot after the acceptance procedure (including screening in the case of a refused lot). Its expectation is the average outgoing quality, AOQ. If it is expressed in terms of the fraction defective, clearly we have

$$\text{AOQ} = pL(p)$$

where p is the original value of the fraction defective and $L(x)$ is the operating characteristic function. The AOQ curve always has a maximum (for high values of p, AOQ will be low since almost all lots will be rejected and screened; see Fig. 53). The maximum of AOQ is called *average outgoing quality limit* (AOQL). This represents the worst possible average quality accepted by the plan whatever the incoming quality may be. Generally, AOQL is between AQL and LQ.

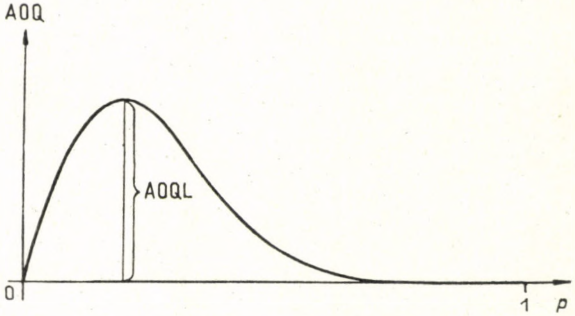


FIG. 53. Average outgoing quality

In the schemes of LQ as well as in those of AOQL, Dodge and Romig base the determination of the sample size on the requirement that the cost of inspection (including sampling and screening) should be minimal if the incoming quality agrees with the "process average" indicated in the Table. This means the schemes are semi-economically based.

In summarizing: the schemes guarantee the value of LQ or AOQL whatever the incoming quality is, and guarantee minimal inspection cost if the hypothesis of process quality ("process average") is true.

In the schemes, the consumer's risk is fixed at 10%. LQ (LTPD) ranges in the tables from 0.5% to 10%; the AOQL from 0.1% to 10%. Volume [C 2] contains single and double sampling plans.

If the schemes are actually used for acceptance (different vendor and consumer), the vendor has to pay all the inspection cost, including screening. The LQ and AOQL are to be chosen by the consumer. The determination of the "process average" may be left in principle to the vendor, since the use of an incorrect process average results in high costs for him and has no effect on the protection given by the scheme to the consumer (the value of the outgoing quality limit being independent of it).

If the assumed conditions are not fulfilled, the use of the schemes may be fairly uneconomical: the incoming quality may be much better than LQ and considerably better than AOQL, and yet a large proportion of the lots may be rejected. Also MIL-STD-105D gives AOQL values in Tables V-A (reproduced in our Table 21) and V-B.

2.2.8 Some Other Approaches to Sampling Inspection by Attributes

The problems of the *posterior distribution* in sampling, under given prior distribution, have been studied by Oderfeld [B 53], Sittig [B 72], Champernown [B 15], Barnard [B 5], Sarkadi [B 65], and James [B 44].

The case of a beta-type prior distribution is particularly advantageous because here the posterior distribution is of the same type (with different parameters, see II.4.2.4.4), and the posterior probability limits of the fraction defective can be computed easily. Another simple case is that of a discrete prior distribution with two possible values (two-point distribution).

The *decision theory approach* to sampling inspection is summarized and discussed in the paper by Wetherill and Campling [B 84]. Different models are considered by Hald [B 34—37], Guthrie and Johns [B 33], and Griffiths and Rao [B 32].

The costs always determine a so-called *breakeven quality*, separating values of fraction defectives for which acception is advantageous and disadvantageous, respectively. Under regular circumstances this point must always lie within the steep section of the operating characteristic curve; its exact optimal position depends on the cost factors and the prior distribution.

Hald [B 34] edited sampling inspection tables for two-point prior distribution and for some other prior distributions as well.

The lot size–sample size ratio in the models of Guthrie and Johns [B 33] and Hald [B 34] makes $n = N^{1/2}$ where N denotes the lot size and n the sample size.

TABLE

(Table V-A of MIL-

AVERAGE OUTGOING QUALITY LIMIT

Code letter	Sample size	Acceptable quality level (%)										
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0
A	2											
B	3											
C	5											
D	8											
E	13											2.8
F	20										1.8	
G	32									1.2		
H	50								0.74			1.7
J	80							0.46			1.1	1.7
K	125						0.29			0.67	1.1	1.6
L	200					0.18			0.42	0.69	0.97	1.6
M	315				0.12			0.27	0.44	0.62	1.00	1.4
N	500			0.074			0.17	0.27	0.39	0.63	0.90	1.3
P	800		0.046			0.11	0.17	0.24	0.40	0.56	0.82	1.2
Q	1250	0.029			0.067	0.11	0.16	0.25	0.36	0.52	0.75	1.2
R	2000			0.042	0.069	0.097	0.16	0.22	0.33	0.47	0.73	

Note. For the exact AOQL, the above values must be multiplied by $\left(1 - \frac{\text{Sample size}}{\text{Lot size}}\right)$.

The "*point of control*", i.e., the quality at which the probability of acceptance is 50%, serves as a basis of the Philips Standard Sampling System.

In some cases, plans connected with *narrow-limit-gages* can be applied; for this, we refer to Section 1.3.4. As the "fraction defective" is here generally higher than usual, the sampling tables cannot be used: Cavé ([A 6], p. 528) gives a graph for this purpose.

If the product is delivered in a continuous stream rather than on a lot-by-lot basis, *continuous sampling tables* may be applied [D 6]. This scheme was first proposed by H. F. Dodge. In this scheme, inspection starts with 100 per cent inspection; if i successive units are found, it is switched to an inspection level in which each kth item will be inspected. Switch back to 100 per cent inspection will be instituted as soon as c out of i consecutive inspected items will be found. The basic principle of the *multilevel continuous sampling scheme* [D 5] is the same but the number of inspection levels is more than 2.

21

STD-105D)

FACTORS FOR NORMAL INSPECTION

Acceptable quality level (%)														
1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
			18			42	69	97	160	220	330	470	730	1100
		12			28	46	65	110	150	220	310	490	720	1100
	7.4			17	27	39	63	90	130	190	290	430	660	
4.6			11	17	24	40	56	82	120	180	270	410		
		6.5	11	15	24	34	50	72	110	170	250			
	4.2	6.9	9.7	16	22	33	47	73						
2.6	4.3	6.1	9.9	14	21	29	46							
2.7	3.9	6.3	9.0	13	19	29								
2.4	4.0	5.6	8.2	12	18									
2.5	3.6	5.2	7.5	12										
2.2	3.3	4.7	7.3											
2.1	3.0	4.7												
1.9	2.9													
1.8														

Values tabulated are read in per cent.

Chain-sampling inspection plans (see, e.g., [A 27], pp. 13/96–97) are similar to double sampling plans, but instead of taking a second sample from the current lot, the results of the last *i* samples from the preceding lots will be taken into account. This scheme was proposed by H. F. Dodge.

2.3 SAMPLING INSPECTION BY VARIABLES

We have already mentioned that in the majority of cases only inspection by attributes is practicable. Sampling by variables requires, first of all, that the critical quality characteristic be a measurable quantity. Difficulties arise if there are more equally important characteristics. In all cases measurement involves higher inspection costs per item than inspection by attributes (gage). An important requirement is some knowledge of the shape of the distribution of the initial variables.

Here we will consider the case where this distribution is normal.

The most widely used sampling scheme by variables is the MIL-STD-414

(a) *One-sided specification, known standard deviation.* Let us first consider the simplest case, where there is a one-sided specification and the population standard deviation is known and can be assumed to be stable. In this case the hypothesis testing problems concerning the population mean and the fraction defective are equivalent i.e., the u test is appropriate. Since in acceptance sampling the fraction defective is important, sampling tables index the plans in terms of this quantity. Values of the distribution mean can be easily converted into values of fraction defective: for example, if there is an upper specification limit U, then the hypothesis that the fraction defective is less or equal to p_0 is equivalent to the hypothesis that the distribution mean is less or equal to $\mu_0 = U + \sigma\Phi^{-1}(p_0)$, where σ denotes the known standard deviation of the distribution and $\Phi^{-1}(x)$ the inverse of the standardized normal distribution function.

TABLE 22

(Table A-2 of MIL-STD-414)

SAMPLE SIZE CODE LETTERS FOR SAMPLING BY VARIABLES

Lot size	Inspection levels				
	I	II	III	IV	V
3 to 8	B	B	B	B	C
9 to 15	B	B	B	B	D
16 to 25	B	B	B	C	E
26 to 40	B	B	B	D	F
41 to 65	B	B	C	E	G
66 to 110	B	B	D	F	H
111 to 180	B	C	E	G	I
181 to 300	B	D	F	H	J
301 to 500	C	E	G	I	K
501 to 800	D	F	H	J	L
801 to 1 300	E	G	I	K	L
1 301 to 3 200	F	H	J	L	M
3 201 to 8 000	G	I	L	M	N
8 001 to 22 000	H	J	M	N	O
22 001 to 110 000	I	K	N	O	P
110 001 to 550 000	I	K	O	P	Q
550 001 & Over	I	K	P	Q	Q

Note. Sample size code letters given in the body of the table are applicable when the indicated inspection levels are to be used.

Let μ be the expected value of the underlying normal distribution. This means that the fraction defective is

$$p = \Phi\left(\frac{\mu - U}{\sigma}\right).$$

Now, p being a monotone increasing function of μ, we have $p \leqslant p_0$ if and only if $\mu \leqslant \mu_0$.

MIL-STD-414 shows similarities with the attributes schemes based on AQL: it has a code letter table (reproduced in our Table 22), and the tables giving the sampling plans are arranged according to the code letter and the AQL. Moreover, MIL-STD-414 has an even closer connection with MIL-STD-105B (predecessor of MIL-STD-105D): plans having the same code letter and AQL have practically the same OC function in both schemes. This means that they are to be considered equivalent with respect to their effect; naturally, the sample size is lower with inspection by variables. In a considerable number of cases, the saving in inspection costs surpasses the surplus cost connected with measurement.

Table D-1 of MIL-STD-414 (part of which is reproduced in Table 23) is the master table for the case of known variance and unilateral specification limit. It gives the sample size (n) and the critical value (k) of the ratio $(\bar{x} - L)/\sigma$ or $(U - \bar{x})/\sigma$. (If $(\bar{x} - L)/\sigma$ or $(U - \bar{x})/\sigma$ is larger than k, then the lot is to be refused.) Here σ is the population standard deviation, L (U) the lower (upper) specification limit, $\bar{x}$ the sample mean.

(b) *One-sided specification, population standard deviation unknown.* In this case the uniformly most powerful invariant test has an acceptance region of the following type ([A 30], p. 224):

$$\frac{\bar{x} - L}{s} \leqslant k$$

or

$$\frac{U - \bar{x}}{s} \geqslant k$$

respectively. Here s denotes the sample standard deviation; k the critical value (it is, of course, not the same function of the sample size and of the level of significance as in the preceding case), U, L and $\bar{x}$ have the same meanings as before. $(U - \bar{x})\sqrt{n}/s$ and $(\bar{x} - L)\sqrt{n}/s$ are distributed according to a non-central t distribution, with $n - 1$ degrees of freedom and non-centrality parameters $(U - \mu)\sqrt{n}/\sigma$ and $(\mu - L)\sqrt{n}/\sigma$, respectively (n is the sample size, μ and σ the population mean and standard de-

TABLE 23

SAMPLING PLANS BASED ON VARIABILITY KNOWN
(From Table D-1 of MIL-STD-414)
(Single specification limit)

Sample size code letter	Acceptable quality levels (normal inspection)													
	.40		.65		1.00		1.50		2.50		4.00		6.50	
	n	k	n	k	n	k	n	k	n	k	n	k	n	k
B	↓		↓		↓		↓		↓		↓		↓	
C	↓		↓		2	1.36	2	1.25	2	1.09	2	.936	3	.755
D	↓		2	1.58	2	1.42	2	1.33	3	1.17	3	1.01	3	.825
E	2	1.81	3	1.69	3	1.56	3	1.44	4	1.28	4	1.11	5	.919
F	3	1.91	4	1.80	4	1.69	4	1.53	5	1.39	5	1.20	6	.991
G	5	2.05	5	1.88	6	1.78	6	1.62	7	1.45	8	1.28	9	1.07
H	6	2.08	7	1.95	7	1.80	8	1.68	9	1.49	10	1.31	12	1.11
I	8	2.13	8	1.96	9	1.83	10	1.70	11	1.51	13	1.34	15	1.13
J	9	2.13	10	1.99	11	1.86	12	1.72	13	1.53	15	1.35	18	1.15
K	10	2.16	11	2.01	12	1.88	14	1.75	15	1.56	18	1.38	20	1.17
L	12	2.18	13	2.03	14	1.89	15	1.75	18	1.57	20	1.38	23	1.17
M	14	2.21	16	2.07	17	1.93	19	1.79	22	1.61	25	1.42	29	1.21
N	21	2.27	23	2.12	25	1.97	28	1.84	32	1.65	36	1.46	42	1.24
O	27	2.29	30	2.14	33	2.00	36	1.86	42	1.67	48	1.48	55	1.26
P	40	2.33	44	2.17	49	2.03	54	1.89	61	1.69	70	1.51	82	1.29
Q	54	2.34	59	2.18	65	2.04	71	1.89	81	1.70	93	1.51	109	1.29
	.65		1.00		1.50		2.50		4.00		6.50		10.00	
	Acceptable quality levels (tightened inspection)													

All AQL values are in percent defective.

↓ Use first sampling plan below arrow, that is sample size as well as k value. When sample size equals or exceeds lot size, every item in the lot must be inspected.

viation). Therefore, for given n, the critical value k can be determined with the aid of the distribution function of the noncentral t distribution [C 10]. The appropriate value of n fulfilling the requirements relating to the OC function can be determined by trial and error. In fact, practical fitting of the OC function with those of MIL-STD-105B has been possible with the value of n determined uniquely by the code letter (i.e., not depending on the AQL; in this respect the case of unknown variance differs from the case of known variance and is similar to the attribute case).

The corresponding values of n and k are given in Table 24 (Table B-1 of MIL-STD-414).

Table 25 (Table C-1 of the Standard) serves the same purpose if the sample range (mean range) is calculated instead of the empirical standard deviation. Here the values of k have been obtained with the help of numerical integration.

The subgroup size is always 5 in these plans, except if the sample size is 10 or less.

(c) *Two-sided specification.* Here we deal only with the problem of testing the overall fraction defective. MIL-STD-414 gives instructions for the case when the simultaneous testing of two alternative hypotheses (concerning the fractions defective due to surpassing the lower and the upper specification limits, respectively) is desirable; in this case, however, the methods described above are to be applied for both sides; therefore, this case does not require further detailed discussion here. For practical details, the reader is referred to the instructions given in the standard.

For the overall fraction defective case, we learn by calculations similar to those used in the unilateral case that the hypothesis that the fraction defective is $\leqslant p$ is equivalent with the hypothesis $L + a\sigma \leqslant \mu \leqq U - a\sigma$, provided that $\Phi\left(\frac{L-U}{2\,\sigma}\right) \leqslant \frac{p}{2}$. Here a is the smaller of the two solutions of the equation

$$\Phi(-a) + \Phi\left(a - \frac{U-L}{\sigma}\right) = p\,;$$

μ, σ, L, U have the same meanings as before. If

$$\Phi\left(\frac{L-U}{2\,\sigma}\right) > \frac{p}{2}$$

then the fraction defective is greater than p for any value of μ.

Table 24

(Table B-1 of MIL-STD-414)

Sampling plans based on variability unknown

(Standard deviation method, single specification limit)

Sample size code letter	Sample size	Acceptable quality levels (normal inspection)													
		.04	.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
		k	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*
B	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	1.12	.958	.765	.566	.341
C	4	↓	↓	↓	↓	↓	↓	↓	1.45	1.34	1.17	1.01	.814	.617	.393
D	5	↓	↓	↓	↓	↓	↓	1.65	1.53	1.40	1.24	1.07	.874	.675	.455
E	7	↓	↓	↓	↓	2.00	1.88	1.75	1.62	1.50	1.33	1.15	.955	.755	.536
F	10	↓	↓	↓	2.24	2.11	1.98	1.84	1.72	1.58	1.41	1.23	1.03	.828	.611
G	15	2.64	2.53	2.42	2.32	2.20	2.06	1.91	1.79	1.65	1.47	1.30	1.09	.886	.664
H	20	2.69	2.58	2.47	2.36	2.24	2.11	1.96	1.82	1.69	1.51	1.33	1.12	.917	.695
I	25	2.72	2.61	2.50	2.40	2.26	2.14	1.98	1.85	1.72	1.53	1.35	1.14	.936	.712
J	30	2.73	2.61	2.51	2.41	2.28	2.15	2.00	1.86	1.73	1.55	1.36	1.15	.946	.723
K	35	2.77	2.65	2.54	2.45	2.31	2.18	2.03	1.89	1.76	1.57	1.39	1.18	.969	.745
L	40	2.77	2.66	2.55	2.44	2.31	2.18	2.03	1.89	1.76	1.58	1.39	1.18	.971	.746
M	50	2.83	2.71	2.60	2.50	2.35	2.22	2.08	1.93	1.80	1.61	1.42	1.21	1.00	.774
N	75	2.90	2.77	2.66	2.55	2.41	2.27	2.12	1.98	1.84	1.65	1.46	1.24	1.03	.804
O	100	2.92	2.80	2.69	2.58	2.43	2.29	2.14	2.00	1.86	1.67	1.48	1.26	1.05	.819
P	150	2.96	2.84	2.73	2.61	2.47	2.33	2.18	2.03	1.89	1.70	1.51	1.29	1.07	.841
Q	200	2.97	2.85	2.73	2.62	2.47	2.33	2.18	2.04	1.89	1.70	1.51	1.29	1.07	.845
		.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00	
		Acceptable quality levels (tightened inspection)													

All AQL values are in percent defective.

↓ Use first sampling plan below arrow, that is, sample size as well as *k* value. When sample size equals or exceeds lot size, every item in the lot must be inspected.

TABLE 25

(Table C-1 of MIL-STD-414)

SAMPLING PLANS BASED ON VARIABILITY UNKNOWN

(Range method, single specification limit)

Sample size code letter	Sample size	Acceptable quality levels (normal inspection)													
		.04	.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
		k	k	k	k	k	k	k	k	k	k	k	k	k	k
B	3								↓	↓	.587	.502	.401	.296	.178
C	4							↓	.651	.598	.525	.450	.364	.276	.176
D	5					↓	↓	.663	.614	.565	.498	.431	.352	.272	.184
E	7				↓	.702	.659	.613	.569	.525	.465	.405	.336	.266	.189
F	10	↓	↓	↓	.916	8.63	.811	.755	.703	.650	.579	.507	.424	.341	.252
G	15	1.09	1.04	.999	.958	.903	.850	.792	.738	.684	.610	.536	.452	.368	.276
H	25	1.14	1.10	1.05	1.01	.951	.896	.835	.779	.723	.647	.571	.484	.398	.305
I	30	1.15	1.10	1.06	1.02	.959	.904	.843	.787	.730	.654	.577	.490	.403	.310
J	35	1.16	1.11	1.07	1.02	.964	.908	.848	.791	.734	.658	.581	.494	.406	.313
K	40	1.18	1.13	1.08	1.04	.978	.921	.860	.803	.746	.668	.591	.503	.415	.321
L	50	1.19	1.14	1.09	1.05	.988	.931	.893	.812	.754	.676	.598	.510	.421	.327
M	60	1.21	1.16	1.11	1.06	1.00	.948	.885	.826	.768	.689	.610	.521	.432	.336
N	85	1.23	1.17	1.13	1.08	1.02	.962	.899	.839	.780	.701	.621	.530	.441	.345
O	115	1.24	1.19	1.14	1.09	1.03	.975	.911	.851	.791	.711	.631	.539	.449	.353
P	175	1.26	1.21	1.16	1.11	1.05	.994	.929	.868	.807	.726	.644	.552	.460	.363
Q	230	1.27	1.21	1.16	1.12	1.06	.996	.931	.870	.809	.728	.646	.553	.462	.364
		.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00	
		Acceptable quality levels (tightened inspection)													

All AQL values are in percent defective.

↓ Use first sampling plan below arrow, that is, sample size as well as k value. When sample size equals or exceeds lot size, every item in the lot must be inspected.

TABLE 26

(Table C-3 of MIL-STD-414)
SAMPLING PLANS BASED ON VARIABILITY UNKNOWN
(Range method, double specification limit)

Sample size code letter	Sample size	c factor	Acceptable quality levels (normal inspection)													
			.04	.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
			M	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*
B	3	1.910	↓	↓	↓	↓	↓	↓	↓	↓	↓	7.59	18.86	26.94	33.69	40.47
C	4	2.234	↓	↓	↓	↓	↓	↓	↓	1.53	5.50	10.92	16.45	22.86	29.45	36.90
D	5	2.474	↓	↓	↓	↓	↓	↓	1.42	3.44	5.93	9.90	14.47	20.27	26.59	33.95
E	7	2.830	↓	↓	↓	↓	.28	.89	1.99	3.46	5.32	8.47	12.35	17.54	23.50	30.66
F	10	2.405	↓	↓	↓	.23	.58	1.14	2.05	3.23	4.77	7.42	10.79	15.49	21.06	27.90
G	15	2.379	.061	.136	.253	.430	.786	1.30	2.10	3.11	4.44	6.76	9.76	14.09	19.30	25.92
H	25	2.358	.125	.214	.336	.506	.827	1.27	1.95	2.82	3.96	5.98	8.65	12.59	17.48	23.79
I	30	2.353	.147	.240	.366	.537	.856	1.29	1.96	2.81	3.92	5.88	8.50	12.36	17.19	23.42
J	35	2.349	.165	.261	.391	.564	.883	1.33	1.98	2.82	3.90	5.85	8.42	12.24	17.03	23.21
K	40	2.346	.160	.252	.375	.539	.842	1.25	1.88	2.69	3.73	5.61	8.11	11.84	16.55	22.38
L	50	2.342	.169	.261	.381	.542	.838	1.25	1.60	2.63	3.64	5.47	7.91	11.57	16.20	22.26
M	60	2.339	.158	.244	.356	.504	.781	1.16	1.74	2.47	3.44	5.17	7.54	11.10	15.64	21.63
N	85	2.335	.156	.242	.350	.493	.755	1.12	1.67	2.37	3.30	4.97	7.27	10.73	15.17	21.05
O	115	2.333	.153	.230	.333	.468	.718	1.06	1.58	2.25	3.14	4.76	6.99	10.37	14.74	20.57
P	175	2.331	.139	.210	.303	.427	.655	.972	1.46	2.08	2.93	4.47	6.60	9.89	14.15	19.88
Q	230	2.330	.142	.215	.308	.432	.661	.976	1.47	2.08	2.92	4.46	6.57	9.84	14.10	19.82
			.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00	
			Acceptable quality levels (tightened inspection)													

All AQL and table values are in percent defective.

↓ Use first sampling plan below arrow, that is, sample size as well as *M* value. When sample size equals or exceeds lot size, every item in the lot must be inspected.

If σ is known, the hypothesis $L + a\sigma \leqslant \mu \leqslant U - a\sigma$ can be tested by means of the acceptance region $L + k\sigma \leqslant \bar{x} \leqslant U - k\sigma$, where k is the unique solution of the equation

$$\Phi\left(\sqrt{n}\left[\frac{U-L}{\sigma} - a - k\right]\right) - \Phi\left(\sqrt{n}[k - a]\right) = 1 - \alpha,$$

α being the size of the test.

This means that now k depends on $(U - L)\sigma$, i.e., it is not uniquely determined by n, α and AQL, in contrast to the case of unilateral specification.

Hence difficulties arise in constructing sampling tables based on the above relations.

Even more complicated is the case of unknown variance. For this, we refer to [A 14], pp. 54–56.

For practical purposes, the approximation used in the Military Standard-414 is suitable. This method is based on the following principle (see [B 47]).

On the basis of the sample, the minimum-variance unbiased estimator of the fraction defective is computed. This value is compared with its critical values; these critical values are determined in such a way that in the unilateral case they define exactly the test described in (a) and in (b). (The approximation lies in the assumption that the OC function is approximately independent of whether the specification is unilateral or bilateral: according to [B 47], numerical calculation supports this idea.)

The minimum-variance unbiased estimator has the additivity property: it can be calculated separately for both sides and the results should be added up.

In the case of unknown variance, Table B-5 of the Standard gives the unilateral estimators. We know $\frac{\bar{x} - L}{s}$ and $\frac{U - \bar{x}}{s}$ and from the Table obtain the unilateral estimators $\hat{p}_l$ and $\hat{p}_u$, and calculate the overall estimate $\hat{p} = \hat{p}_u + \hat{p}_l$. The critical values M of $\hat{p}$ are given in Table B-3 of the Standard.

The procedure is essentially the same in the case of known variance and in the range method of unknown variance case. (In the latter method, the estimators are no longer minimum-variance unbiased estimators.) The corresponding Tables in the Standard are D-5 and D-3 in the case of known variance and C-5 (see our Table 27) and C-3 (our Table 26) in

TABLE

TABLE FOR ESTIMATING THE
(From Table C-5
(Variability unknown,

Q_U or Q_L	Sample size 10	Sample size 15	Sample size 25
.0	50.00	50.00	50.00
.1	46.20	46.13	46.08
.2	42.42	42.29	42.19
.3	38.70	38.51	38.38
.31	38.33	38.14	38.00
.32	37,96	37.77	37.63
.33	37.60	37.39	37.25
.34	37.23	37.02	36.88
.35	36.87	36.65	36.50
.36	36.50	36.29	36.13
.37	36.14	35.92	35.76
.38	35.78	35.55	35.39
.39	35.41	35.19	35.02
.40	35.05	34.82	34.66
.41	34.69	34.46	34.29
.42	34.33	34.10	33.93
.43	33.93	33.74	33.57
.44	33.62	33.38	33.21
.45	33.27	33.02	32.85
.46	32.91	32.66	32.49
.47	32.56	32.31	32.13
.48	32.21	31.96	31.78
.49	31.86	31.60	31.42
.50	31.51	31.25	31.07
.51	31.16	30.90	30.72
.52	30.81	30.55	30.37
.53	30.46	30.21	30.02
.54	30.12	29.86	29.68
.55	29.78	29.52	29.33
.56	29.44	29.18	28.99
.57	29.09	28.83	28.65
.58	28.76	28.50	28.31
.59	28.42	28.16	27.97
.60	28.08	27.82	27.64
.61	27.75	27.49	27.31
.62	27.41	27.16	26.97
.63	27.08	26.82	26.64
.64	26.75	26.50	26.32
.65	26.42	26.17	25.99
.66	26.10	25.84	25.67
.67	25.77	25.52	25.34
.68	25.45	25.20	25.02
.69	25.12	24.88	24.71
.70	24.80	24.56	24.39
.71	24.48	24.24	24.07
.72	24.17	23.93	23.76
.73	23.85	23.61	23.45
.74	23.54	23.30	23.14
.75	23.22	22.99	22.84
.76	22.91	22.69	22.53
.77	22.60	22.38	22.23
.78	22.30	22.08	21.93
.79	21.99	21.78	21.64
.80	21.69	21.48	21.34
.81	21.39	21.18	21.04
.82	21.09	20.89	20.75
.83	20.79	20.59	20.46
.84	20.49	20.30	20.17
.85	20.20	20.01	19.89
.86	19.90	19.73	19.60
.87	19.61	19.44	19.32
.88	19.33	19.16	19.04
.89	19.04	18.88	18.77
.90	18.75	18.60	18.50
.91	18.47	18.32	18.22
.92	18.19	18.05	17.96
.93	17.91	17.78	17.69
.94	17.64	17.51	17.43
.95	17.36	17.24	17.17
.96	17.09	16.98	16.91
.97	16.82	16.71	16.65
.98	16.55	16.45	16.39
.99	16.28	16.19	16.14
1.00	16.02	15.94	15.89
1.01	15.76	15.68	15.64
1.02	15.50	15.43	15.40
1.03	15.24	15.18	15.15
1.04	14.98	14.94	14.91
1.05	14.73	14.69	14.67
1.06	14.48	14.45	14.44
1.07	14.23	14.21	14.20
1.08	13.98	13.97	13.97
1.09	13.74	13.73	13.74

$Q_U = \frac{(U - \bar{x})c}{\bar{R}}$,

Note. Values tabulated

27

LOT FRACTION DEFECTIVE
of MIL-STD-414)
range method)

Q_U or Q_L	Sample size			Q_U or Q_L	Sample size		
	10	15	25		10	15	25
1.10	13.49	13.50	13.52	1.50	5.66	6.05	6.33
1.11	13.25	13.27	13.29	1.51	5.51	5.91	6.19
1.12	13.02	13.04	13.07	1.52	5.37	5.77	6.06
1.13	12.78	12.81	12.85	1.53	5.22	5.64	5.93
1.14	12.55	12.59	12.63	1.54	5.08	5.50	5.80
1.15	12.31	12.37	12.42	1.55	4.94	5.37	5.68
1.16	12.08	12.15	12.21	1.56	4.81	5.24	5.55
1.17	11.86	11.93	12.00	1.57	4.67	5.11	5.43
1.18	11.63	11.71	11.79	1.58	4.54	4.99	5.31
1.19	11.41	11.50	11.58	1.59	1.41	4.86	5.19
1.20	11.19	11.29	11.38	1.60	4.28	4.74	5.08
1.21	10.97	11.08	11.18	1.61	4.16	4.62	4.96
1.22	10.76	10.88	10.98	1.62	4.03	4.51	4.85
1.23	10.54	10.67	10.78	1.63	3.91	4.39	4.74
1.24	10.33	10.47	10.58	1.64	3.79	4.28	4.63
1.25	10.12	10.27	10.39	1.65	3.68	4.17	4.52
1.26	9.92	10.08	10.20	1.66	3.56	4.06	4.41
1.27	9.71	9.88	10.01	1.67	3.45	3.95	4.31
1.28	9.51	9.69	9.83	1.68	3.34	3.85	4.21
1.29	9.31	9.50	9.64	1.69	3.23	3.74	4.10
1.30	9.11	9.32	9.47	1.70	3.13	3.64	4.00
1.31	8.92	9.13	9.29	1.71	3.02	3.54	3.92
1.32	8.73	8.95	9.11	1.72	2.92	3.45	3.82
1.33	8.54	8.77	8.94	1.73	2.82	3.35	3.73
1.34	8.35	8.59	8.77	1.74	2.73	3.26	3.63
1.35	8.16	8.41	8.60	1.75	2.63	3.16	3.54
1.36	7.98	8.24	8.43	1.76	2.54	3.07	3.45
1.37	7.80	8.07	8.27	1.77	2.45	2.99	3.37
1.38	7.62	7.90	8.11	1.78	2.36	2.90	3.28
1.39	7.45	7.73	7.95	1.79	2.27	2.81	3.20
1.40	7.27	7.57	7.79	1.80	2.19	2.73	3.11
1.41	7.10	7.41	7.63	1.81	2.10	2.65	3.03
1.42	6.93	7.25	7.48	1.82	2.02	2.57	2.96
1.43	6.76	7.09	7.33	1.83	1.94	2.49	2.88
1.44	6.60	6.93	7.18	1.84	1.87	2.42	2.80
1.45	6.44	6.78	7.03	1.85	1.79	2.34	2.73
1.46	6.28	6.63	6.89	1.86	1.72	2.27	2.66
1.47	6.12	6.48	6.74	1.87	1.65	2.20	2.59
1.48	5.96	6.34	6.60	1.88	1.58	2.13	2.52
1.49	5.81	6.19	6.47	1.89	1.51	2.06	2.45

$Q_L = \dfrac{(\bar{x} - L)c}{\bar{R}}$.
are read in percent.

TABLE 27

Q_U or Q_L	Sample size			Q_U or Q_L	Sample size		
	10	15	25		10	15	25
1.90	1.45	1.99	2.38	2.30	0.089	0.386	0.672
1.91	1.38	1.93	2.32	2.31	0.079	0.367	0.648
1.92	1.32	1.86	2.25	2.32	0.070	0.348	0.624
1.93	1.26	1.80	2.19	2.33	0.061	0.330	0.601
1.94	1.20	1.74	2.13	2.34	0.054	0.313	0.579
1.95	1.15	1.68	2.07	2.35	0.047	0.296	0.558
1.96	1.09	1.62	2.01	2.36	0.040	0.280	0.538
1.97	1.04	1.57	1.95	2.37	0.035	0.265	0.518
1.98	0.99	1.51	1.90	2.38	0.029	0.250	0.498
1.99	0.94	1.46	1.84	2.39	0.025	0.236	0.479
2.00	0.89	1.41	1.79	2.40	0.021	0.223	0.461
2.01	0.84	1.36	1.74	2.41	0.017	0.210	0.443
2.02	0.80	1.31	1.69	2.42	0.014	0.198	0.426
2.03	0.75	1.26	1.64	2.43	0.011	0.186	0.410
2.04	0.71	1.21	1.59	2.44	0.009	0.175	0.393
2.05	0.67	1.17	1.54	2.45	0.007	0.165	0.378
2.06	0.63	1.12	1.49	2.46	0.005	0.154	0.362
2.07	0.60	1.08	1.45	2.47	0.004	0.145	0.348
2.08	0.56	1.04	1.40	2.48	0.003	0.136	0.333
2.09	0.53	1.00	1.36	2.49	0.002	0.127	0.321
2.10	0.49	0.96	1.32	2.50	0.001	0.118	0.307
2.11	0.46	0.92	1.28	2.51	0.001	0.111	0.294
2.12	0.43	0.88	1.24	2.52	0.000	0.103	0.282
2.13	0.40	0.85	1.20	2.53	0.000	0.096	0.270
2.14	0.38	0.81	1.16	2.54	0.000	0.089	0.258
2.15	0.35	0.78	1.13	2.55	0.000	0.083	0.247
2.16	0.32	0.75	1.09	2.56	0.000	0.077	0.237
2.17	0.30	0.71	1.06	2.57	0.000	0.071	0.227
2.18	0.28	0.68	1.02	2.58	0.000	0.066	0.217
2.19	0.26	0.65	0.99	2.59	0.000	0.061	0.207
2.20	0.236	0.625	0.954	2.60	0.000	0.056	0.196
2.21	0.217	0.597	0.922	2.61	0.000	0.052	0.189
2.22	0.199	0.570	0.891	2.62	0.000	0.048	0.181
2.23	0.182	0.544	0.861	2.63	0.000	0.044	0.173
2.24	0.166	0.519	0.831	2.64	0.000	0.040	0.164
2.25	0.150	0.495	0.802	2.65	0.000	0.037	0.157
2.26	0.136	0.471	0.775	2.66	0.000	0.034	0.149
2.27	0.123	0.449	0.748	2.67	0.000	0.031	0.143
2.28	0.111	0.427	0.722	2.68	0.000	0.028	0.136
2.29	0.099	0.406	0.697	2.69	0.000	0.025	0.129

(cont.)

Q_U or Q_L	Sample size			Q_U or Q_L	Sample size		
	10	15	25		10	15	25
2.70	0.000	0.023	0.123	3.10	0.000	0.000	0.011
2.71	0.000	0.021	0.117	3.11	0.000	0.000	0.011
2.72	0.000	0.019	0.111	3.12	0.000	0.000	0.010
2.73	0.000	0.017	0.106	3.13	0.000	0.000	0.010
2.74	0.000	0.015	0.101	3.14	0.000	0.000	0.009
2.75	0.000	0.014	0.096	3.15	0.000	0.000	0.008
2.76	0.000	0.012	0.091	3.16	0.000	0.000	0.008
2.77	0.000	0.011	0.086	3.17	0.000	0.000	0.007
2.78	0.000	0.010	0.081	3.18	0.000	0.000	0.006
2.79	0.000	0.008	0.077	3.19	0.000	0.000	0.006
2.80	0.000	0.007	0.074	3.20	0.000	0.000	0.005
2.81	0.000	0.007	0.070	3.21	0.000	0.000	0.005
2.82	0.000	0.006	0.066	3.22	0.000	0.000	0.004
2.83	0.000	0.005	0.062	3.23	0.000	0.000	0.004
2.84	0.000	0.004	0.059	3.24	0.000	0.000	0.004
2.85	0.000	0.004	0.055	3.25	0.000	0.000	0.003
2.86	0.000	0.003	0.053	3.26	0.000	0.000	0.003
2.87	0.000	0.003	0.050	3.27	0.000	0.000	0.003
2.88	0.000	0.002	0.047	3.28	0.000	0.000	0.003
2.89	0.000	0.002	0.044	3.29	0.000	0.000	0.003
2.90	0.000	0.002	0.042	3.30	0.000	0.000	0.003
2.91	0.000	0.001	0.039	3.31	0.000	0.000	0.003
2.92	0.000	0.001	0.037	3.32	0.000	0.000	0.002
2.93	0.000	0.001	0.035	3.33	0.000	0.000	0.002
2.94	0.000	0.001	0.033	3.34	0.000	0.000	0.002
2.95	0.000	0.001	0.031	3.35	0.000	0.000	0.002
2.96	0.000	0.001	0.029	3.36	0.000	0.000	0.002
2.97	0.000	0.000	0.027	3.37	0.000	0.000	0.002
2.98	0.000	0.000	0.025	3.38	0.000	0.000	0.001
2.99	0.000	0.000	0.024	3.39	0.000	0.000	0.001
3.00	0.000	0.000	0.022	3.40	0.000	0.000	0.001
3.01	0.000	0.000	0.022	3.41	0.000	0.000	0.001
3.02	0.000	0.000	0.020	3.42	0.000	0.000	0.001
3.03	0.000	0.000	0.019	3.43	0.000	0.000	0.001
3.04	0.000	0.000	0.017	3.44	0.000	0.000	0.001
3.05	0.000	0.000	0.016	3.45	0.000	0.000	0.001
3.06	0.000	0.000	0.015	3.46	0.000	0.000	0.001
3.07	0.000	0.000	0.014	3.47	0.000	0.000	0.001
3.08	0.000	0.000	0.013	3.48	0.000	0.000	0.001
3.09	0.000	0.000	0.012	3.49	0.000	0.000	0.000

the range method. A difference in both cases is that Tables D-3 and C-3 give factors c with which $\frac{\bar{x}-L}{\sigma}$, $\frac{U-\bar{x}}{\sigma}$, $\frac{\bar{x}-L}{\bar{R}}$ or $\frac{U-\bar{x}}{\bar{R}}$ are to be multiplied before using Table D-5 or C-5.

2.4 SEQUENTIAL SAMPLING

2.4.1 On the Sequential Procedure

The sequential sampling procedure is a method created for quick and economical but reliable quality control of lots. After certain preliminaries, M. Friedman and W. A. Wallis called the attention of A. Wald to this possibility. A. Wald has elaborated the fundamentals of this method in his book "Sequential Analysis". The exact determination of the characteristics of the procedure takes much effort even for simple distribution functions, and in most of the cases only approximate methods exist. Wald has given, however, some very ingenious approximations which are sufficient in the majority of practical applications. We turn to a more detailed treatment of these simple methods and also mention some more rigorous inequalities. As important applications, the cases of binomial, normal and exponential distributions will be given.

In the sequential procedure the sample elements are examined successively (individually or in small groups). Hence this method is suitable, in general, when the examination of a single sample element requires only a short time. It can be adopted, however, for life length investigation as well.

In a sequential test the usual starting point is the testing of a simple null hypothesis against a simple alternative, by means of the probability ratio test for which the Neyman–Pearson lemma (see Chapter II, Section 4.3.4) guarantees the optimality.

This method often provides us with a simple statistical test procedure to be carried out step-by-step. At each step there are three possibilities: to accept the null hypothesis, to accept the alternative, or to go on with the examination of a further sample element. This means that the sample size is not at all restricted: it is a random variable which can take on arbitrarily large values. An interesting feature of the sequential test is that this lack of restriction in the sample size leads to an optimal average sample size, which, in certain cases, is only 40–60% of the corresponding simple sample size. This holds when either the null hypothesis or the alternative hypothesis is true, but there are reassuring results for the intermediate cases as well, which will be considered also.

In the treatment to follow, the general theory will be dealt with first; afterwards we will turn to the applications of the three mentioned types of distributions.

2.4.2 The Sequential Version of the Probability Ratio Test

Let X be a discrete or continuous random variable. If X is discrete the probability of the value $X = x$ will be denoted by $f(x; \vartheta)$; if X is continuous, then $f(x; \vartheta)$ is the density function. This convention is due to Wald and does not cause any difficulty. Hence, for discrete variables,

$$\sum_{(i)} f(x_i;\ \vartheta) = 1$$

holds, where $x_1, x_2, \ldots$ are the possible values of X, while for continuous random variables the relation

$$\int_{-\infty}^{\infty} f(x; \vartheta) d\vartheta = 1$$

is valid.

Concerning the nature of the unknown parameter ϑ, no assumption is needed, but we may start with testing the hypothesis

$$H_0\colon\ \vartheta = \vartheta_0$$

against the alternative

$$H_1\colon\ \vartheta = \vartheta_1.$$

The problem is to be decided at level $1 - \alpha$ and with an error probability β of the second kind.

Let us assume that we have an infinite sequence of observable, independent random variables $X_1, X_2, \ldots$; i.e., a sample of infinite size with common distribution defined by $f(x; \vartheta)$. The number of X_i's actually observed, will depend on chance; hence the number of observed sample elements is a random variable which will be denoted throughout by n. Later on we will be interested in the expected value $E(n \mid \vartheta)$ of n.

The sequential sampling procedure itself is to be carried out in the following way: let us suppose that we have already observed m random variables, i.e., we have taken up to now a sample of size m, resulting in the values $X_1 = x_1$, $X_2 = x_2, \ldots X_m = x_m$. Now we compute the ratio

$$\frac{p_{1m}}{p_{0m}} = \frac{f(x_1; \vartheta_1)\, f(x_2; \vartheta_1) \ldots f(x_m; \vartheta_1)}{f(x_1; \vartheta_0)\, f(x_2; \vartheta_0) \ldots f(x_m; \vartheta_0)}$$

and compare it with two values, $A = A(\alpha, \beta)$ and $B = B(\alpha, \beta)$; $A > B$ (the choice of A and B will be treated below).

If $\frac{p_{1m}}{p_{0m}} \leqslant B$, then H_0 will be accepted;

If $\frac{p_{1m}}{p_{0m}} \geqslant A$, then H_1 will be accepted;

If $B < \frac{p_{1m}}{p_{0m}} < A$, then X_{m+1} will be observed.

This means that in the first two cases the procedure comes to an end, while in the third case it will be continued. This procedure may lead—in principle—to an infinitely large sample, but Wald has proved that with probability one *the sequential probability ratio test comes to an end in a finite number of steps.*

The above procedure has the following simple meaning: In the discrete case, p_{0m} is the probability of the actually observed sample when H_0 is true, while p_{1m} is the probability of this outcome when H_1 is valid. As is known, in mathematical statistics we usually prefer to accept the occurrence of an event which has a high probability than one with a small probability. Therefore we prefer the hypothesis belonging to the higher probability. However, the sequential procedure does not terminate if there is only a slight difference between these two probabilities. If, however, e.g. $\frac{p_{1m}}{p_{0m}} \sim 20$, then the probability of the actual sample is 20 times more under H_1 than under H_0; in such a case the acceptance of H_1 can be taken as justified. For continuous random variables, similar arguments hold, since the density is proportional to the probability that the sample lies in a small domain of the Euclidean m space around the point $(x_1, x_2, \ldots, x_m)$.

But how do we choose the values of A and B? As already indicated by the notation, these quantities depend on the error probabilities of both kinds. The choice

$$B \sim \frac{\beta}{1-\alpha} \quad \text{and} \quad A \sim \frac{1-\beta}{\alpha}$$

is suitable for many practical purposes and a more exact value is not needed. (If, e.g., $\alpha = \beta = 0.05$, then $B \sim 0.053$ and $A \sim 21$. Obviously, in all practical cases $B < 1 < A$ holds.) This can be justified in the following way: Suppose that instead of the exact values $A = A(\alpha, \beta)$ and $B = B(\alpha, \beta)$, the constants $\frac{\beta}{1-\alpha}$ and $\frac{1-\beta}{\alpha}$ are chosen. To these constants there belongs some other sequential sampling procedure, with error probabilities ω and ω', which differ from α and β. We shall prove that

$$\omega \leqslant \frac{\alpha}{1-\beta}, \ \omega' \leqslant \frac{\beta}{1-\alpha}, \quad \omega + \omega' \leqslant \alpha + \beta. \tag{5}$$

This means that the exact theoretical constants A and B are replaced by $\frac{\beta}{1-\alpha}$ and $\frac{1-\beta}{\alpha}$ respectively. This results in only a very moderate increase in the error probabilities.

We turn now to the proof of the first inequality. This is a consequence of the definition. First of all, the critical region of the test has the following form,

$$\mathfrak{X}_c = \left\{ (x_1, x_2, \ldots): \quad B < \frac{p_{1m}}{p_{0m}} < A, \quad m = 1, 2, \ldots n-1, \right.$$
$$\left. \text{but} \frac{p_{1n}}{p_{0n}} \geqslant A; \quad n = 1, 2, \ldots \right\}.$$

This means that for the points of the critical region—for each dimension n— we have $p_{1n} \geqslant Ap_{0n}$, or in other words: the probability (or density) of each critical point is at least A times larger under H_1 than under H_0. But the probability of the totality of the critical points, i.e., the probability of the critical region under H_1, is the power $1 - \beta$ of the test, while under H_0 it is the size α of the test. Consequently, $(1-\beta) \geqslant A\alpha$ holds. Similarly, the inequality $B(1 - \alpha) \geqslant \beta$ can be proved. Let us turn to the modified test procedure with the critical values $\frac{1-\beta}{\alpha}$ and $\frac{\beta}{1-\alpha}$ and the error probabilities ω and ω'. Our argument holds for this sequential procedure as well, from which the corresponding inequalities follow:

$$1 - \omega' \geqslant \frac{1-\beta}{\alpha}\omega \text{ and } \frac{\beta}{1-\alpha}(1-\omega) \geqslant \omega'$$

or

$$\frac{1-\omega'}{\omega} \geqslant \frac{1-\beta}{\alpha} \text{and } \frac{\omega'}{1-\omega} \leqslant \frac{\beta}{1-\alpha}.$$

From the latter system the three relations under (5) can be easily obtained.

We call attention to another point: If ω or ω' were much less than α or β, then this would cause the procedure to be more rigorous than the exact one and would hence require a significantly larger average sample size. It can be proved, however, that this case does not occur either.

We now have a simple device for a sampling procedure which protects us against false decisions if H_0 or H_1 is true. But in many practical situations the parameter ϑ (e.g., fraction defective, length or other measure) may also take on values other than ϑ_0 and ϑ_1. For this case the corresponding probabilities are to be determined as well. This will be done in the next Section.

2.4.3 The OC Function of the Sequential Probability Ratio Test

Let $\vartheta(-\infty < \vartheta < \infty)$ be a real parameter occurring in the probability (density) $f(x;\, \vartheta)$ of the random variable X. Concerning the null hypothesis,

$$H_0\colon\ \vartheta = \vartheta_0$$

and the alternative

$$H_1\colon\ \vartheta = \vartheta_1,$$

the sequential probability ratio test with given constants will be applied; we shall denote this test procedure by

$$S = S\{\vartheta_0, \vartheta_1;\ A = A(\alpha, \beta),\ B = B(\alpha, \beta)\}.$$

By means of an elegant argument the following approximate expression can be obtained for the operating characteristic function of the test S:

$$L(\vartheta) \sim \frac{A^{h(\vartheta)} - 1}{A^{h(\vartheta)} - B^{h(\vartheta)}}, \quad (-\infty < \vartheta < \infty), \tag{6}$$

where $h(\vartheta)$ is the unique real number differing from zero and fulfilling the relation:

$$\int_{-\infty}^{\infty} \left(\frac{f(x;\, \vartheta_1)}{f(x;\, \vartheta_0)}\right)^{h(\vartheta)} f(x;\, \vartheta)\, dx = 1. \tag{7}$$

In the case of discrete distributions, the integral is to be replaced by summation.

The existence and uniqueness of $h(\vartheta)$ will not be considered here; for the proof see, e.g., the book by Wald [A 48], p. 158. One may see, however, that $h(\vartheta_0) = +1$ and $h(\vartheta_1) = -1$ hold.

We now turn to the justification of our statement (6). For a given value of ϑ differing from both ϑ_0 and ϑ_1, the following density function will be considered:

$$f^*(x;\vartheta) = \left(\frac{f(x;\vartheta_1)}{f(x;\vartheta_0)}\right)^{h(\vartheta)} f(x;\vartheta).$$

We shall assume in the following that for this ϑ it is true that $h(\vartheta) > 0$; a similar method can be applied for ϑ if $h(\vartheta) < 0$.

The fact that $f^*(x, \vartheta)$ is a density function (or probability distribution in the discrete case) follows from its nonnegativity and from relation (7).

We now introduce the following hypotheses:

$$H_0^*\colon\ P(X \leqslant x) = \int_{-\infty}^{\infty} f(t;\vartheta)dt,$$

$$H_1^*\colon\ P(X \leqslant x) = \int_{-\infty}^{\infty} f^*(t;\vartheta)dt.$$

Let us denote by

$$S^* = S^*\{f(x;\vartheta),\ f^*(x;\vartheta),\ A^{h(\vartheta)}(\omega,\omega'),\ B^{h(\vartheta)}(\omega,\omega')\}$$

the appropriate sequential probability ratio test for deciding between H_0^* and H_1^*. According to our definition, if the procedure has not terminated in the first $m - 1$ steps, then after the mth step it will be continued if

$$B^{h(\vartheta)} < \frac{f^*(x_1;\vartheta)\, f^*(x_2;\vartheta) \ldots f^*(x_k,\vartheta)}{f(x_1;\vartheta)\, f(x_2;\vartheta) \ldots f(x_k,\vartheta)} < A^{h(\vartheta)},$$

or $k = 1, 2, \ldots, m$. It will come to an end with the acceptance of H_0^* if

$$\frac{f^*(x_1;\vartheta)\, f^*(x_2\,\vartheta) \ldots f^*(x_m;\vartheta)}{f(x_1;\vartheta)\, f(x_2;\vartheta) \ldots f(x_m;\vartheta)} \leqslant B^{h(\vartheta)},$$

and with the acceptance of H_1^* if

$$A^{h(\vartheta)} \leqslant \frac{f^*(x_1;\vartheta)\, f^*(x_2;\vartheta) \ldots f^*(x_m;\vartheta)}{f(x_1;\vartheta)\, f(x_2;\vartheta) \ldots f(x_m;\vartheta)}.$$

Since, according to our assumption $h(\vartheta) > 0$ and

$$\frac{f^*(x;\vartheta)}{f(x;\vartheta)} = \left(\frac{f(x;\vartheta_1)}{f(x;\vartheta_0)}\right)^{h(\vartheta)},$$

the above conditions are equivalent to the relations

$$B < \frac{p_{1k}}{p_{0k}} < A, \quad \frac{p_{1m}}{p_{0m}} \leqslant B, \quad A \leqslant \frac{p_{1m}}{p_{0m}},$$

where p_{0m} and p_{1m} have the same meanings as in Chapter III, Section 2.4.2. These relations have been used for deciding between H_0 and H_1.

This means that if by S^* the null hypothesis H^* will be accepted, then S leads at the same time to the acceptance of H_0. But this is equivalent to the following statement: The probability $L(\vartheta)$ that H_0 will be accepted if the true value of the parameter is ϑ agrees with the probability that S^* leads to the acceptance of H_0^* if X is distributed according to $f^*(x; \vartheta)$.

The probability of the latter event can be determined from the following two approximate relations derived in Section 2.4.2:

$$A^{h(\vartheta)} \sim \frac{1-\omega'}{\omega},$$

$$B^{h(\vartheta)} \sim \frac{\omega'}{1-\omega},$$

both of which are valid for the S^* test procedure. From these there follows

$$\omega \sim \frac{1-B^{h(\vartheta)}}{A^{h(\vartheta)}-B^{h(\vartheta)}}.$$

On the other hand $L(\vartheta) = 1 - \omega$, and hence we obtain finally

$$L(\vartheta) \sim \frac{A^{h(\vartheta)}-1}{A^{h(\vartheta)}-B^{h(\vartheta)}}.$$

Wald has given the following lower and upper limits for $L(\vartheta)$: if $h(\vartheta) > 0$, then

$$\frac{A^{h(\vartheta)}-1}{A^{h(\vartheta)}-\eta_\vartheta B^{h(\vartheta)}} \leqslant L(\vartheta) \leqslant \frac{\delta_\vartheta A^{h(\vartheta)}-1}{\delta_\vartheta A^{h(\vartheta)}-B^{h(\vartheta)}},$$

where η_ϑ is the greatest lower bound of the function

$$\eta_\vartheta(u) = u\,\frac{\displaystyle\int_{-\infty}^{\frac{1}{h(\vartheta)}\log u} e^{h(\vartheta)x} f(x;\vartheta)\,dx}{\displaystyle\int_{-\infty}^{\frac{1}{h(\vartheta)}\log u} f(x;\vartheta)\,dx}, \quad u > 1$$

while δ_ϑ is the least upper bound of the function

$$\frac{\displaystyle\int_{\frac{1}{h(\vartheta)}\log\frac{1}{v}}^{\infty} e^{h(\vartheta)x} f(x;\vartheta)\,dx}{\displaystyle\int_{\frac{1}{h(\vartheta)}\log\frac{1}{v}}^{\infty} f(x;\vartheta)\,dx}, \quad 0 < v < 1.$$

In the case when $h(\vartheta) < 0$, with the above defined constants η_ϑ and δ_ϑ, the following inequality holds:

$$\frac{1 - A^{h(\vartheta)}}{\delta_\vartheta B^{h(\vartheta)} - A^{h(\vartheta)}} \leqslant L(\vartheta) \leqslant \frac{1 - \eta_\vartheta A^{h(\vartheta)}}{B^{h(\vartheta)} - \eta_\vartheta A^{h(\vartheta)}} .$$

The corresponding limits will be determined for the binomial and normal distribution.

2.4.4 The Expected Value of the Sample Size

The notation n has been introduced for the total number of sample elements examined in the sequential probability ratio test.

This means that for $m = 1, 2, \ldots n - 1$ the relations

$$B < \frac{p_{1m}}{p_{0m}} < A$$

hold, but either

$$\frac{p_{1n}}{p_{0n}} \leqslant B \quad \text{or} \quad \frac{p_{1n}}{p_{0n}} \geqslant A$$

takes place. We are interested in the expected value of n when ϑ is the true value of the parameter $E(n \mid \vartheta)$. For an easier treatment, we introduce the random variable.

$$Z = \log \frac{f(X; \vartheta_1)}{f(X; \vartheta_0)}.$$

With this, the relation

$$\log \frac{p_{1m}}{p_{0m}} = \sum_{i=1}^{m} \log \frac{f(x_i; \vartheta_1)}{f(x_i; \vartheta_0)} = z_1 + z_2 + \ldots + z_m$$

holds, where z_i is the observed value of Z_i when the ith experiment results in the value $X_i = x_i$. By means of this notation the sequential procedure can be described in the following way: The test is continued after the mth step, if

$$\log B < z_1 + z_2 + \ldots z_m < \log A,$$

while it terminates if the outcome is either

$$z_1 + z_2 + \ldots + z_m \leqslant \log B \text{ or } z_1 + z_2 + \ldots z_m \leqslant \log A.$$

Denoting the expected value of Z in terms of the parameter ϑ by $E(Z|\vartheta)$, the following formula for the expected value of n is valid:

$$E(n\,|\,\vartheta) = \frac{E(Z_1 + Z_2 + \ldots + Z_n\,|\,\vartheta)}{E(Z\,|\,\vartheta)}\ ,$$

if $E(Z\,|\,\vartheta) \neq 0$. The proof will be omitted here.

If there is a value ϑ^* of ϑ for which the relation $E(Z\,|\,\vartheta^*) = 0$ holds, the expression above will take on the form

$$E(n\,|\,\vartheta^*) = \frac{E[(Z_1 + Z_2 + \ldots Z_n)^2\,|\,\vartheta^*]}{E(Z^2\,|\,\vartheta^*)}\ .$$

A rough approximation to $E(n\,|\,\vartheta)$ can be obtained in the following way: When the test ends, i.e., for the first n for which $\sum_{i=1}^{n} Z_i$ is either $\leqslant \log B$ or $\geqslant \log A$, the excess of this sum over $\log A$ or under $\log B$ will be neglected; under this assumption we can say that a decision occurs when $\sum_{i=1}^{n} Z_i$ is equal to either $\log B$ or $\log A$. The former event occurs with probability $L(\vartheta)$, the latter with probability $1 - L(\vartheta)$; consequently, the approximate expected value of $\sum_{i=1}^{n} Z_i$ at termination has the simple form $L(\vartheta) \log B + (1 - L(\vartheta)) \log A$. Hence, for the average sample number the following relation is valid:

$$E(n\,|\,\vartheta) \sim \frac{L(\vartheta)\ \log\ B + (1 - L(\vartheta))\ \log\ A}{E(Z\,|\,\vartheta)}\ ,$$

if the denominator differs from zero.

Finally, we mention the following optimality property of the sequential probability ratio test, which is due to Wald: *Among all possible tests with an error of the first kind, the probability of which does not exceed* α, *and an error of the second kind, the probability of which is not larger than* β, *and for which the expected sample number under* ϑ_0 *and* ϑ_1 *is finite, the sequential probability ratio test minimizes both* $E(n\,|\,\vartheta_0)$ *and* $E(n\,|\,\vartheta_1)$.

2.4.5 Sequential Test in the Case of a Binomial Distribution

Suppose that a lot of products is to be accepted. A lot is considered acceptable if its fraction defective does not exceed the value $p_0 (0 < p_0 < 1)$. In this case its acceptance with high probability, say $1 - \alpha$, is preferable, while if the fraction defective exceeds the value $p_1 (p_0 < p_1 < 1)$, its acceptance should have a small probability, say β. The acceptance procedure is to be carried out with the aid of the sequential probability ratio test

$$S = S\{p_0, p_1, A = A(\alpha, \beta), B = B(\alpha, \beta)\}.$$

Let the random variable X be defined for a piece chosen randomly from the lot in the following way:

$$X = \begin{cases} 1, & \text{if the piece is defective,} \\ 0, & \text{if the piece is a good one.} \end{cases}$$

The distribution of X is given by

$$\begin{aligned} P(X = 1 \mid p) &= f(1, p) = 1 - p, \\ & & 0 \leqslant p \leqslant 1. \\ P(X = 0 \mid p) &= f(0; p) = p, \end{aligned}$$

For the ith piece examined in the course of the sequential procedure, the relation

$$Z_i = \log \frac{f(X_i, p_1)}{f(X_i; p_0)} = \begin{cases} \log \dfrac{p_1}{p_0}, & \text{if } X_i = 1, \\ \log \dfrac{1 - p_1}{1 - p_0} & \text{if } X_i = 0 \end{cases}$$

holds.

Let us denote by k_m the number of defective items among the first m pieces; i.e., $m - k_m$ is the number of good pieces. Denoting the observed value of Z_i by z_i, after the mth step the sum of the z_i's has the form

$$z_1 + z_2 + \ldots + z_m = k_m \log \frac{p_1}{p_0} + (m - k_m) \log \frac{1 - p_1}{1 - p_0}.$$

Now the sequential testing procedure deals at each step with the following three possibilities:

The hypothesis H_0: $p = p_0$ will be accepted after the mth step, if

$$k_m \log \frac{p_1}{p_0} + (m - k_m) \log \frac{1 - p_1}{1 - p_0} \leqslant \log B;$$

it will be rejected, if

$$k_m \log \frac{p_1}{p_0} + (m - k_m) \log \frac{1 - p_1}{1 - p_0} \geqslant \log A.$$

A further sample element will be taken and examined if

$$\log B < k_m \log \frac{p_1}{p_0} + (m - k_m) \log \frac{1 - p_1}{1 - p_0} < \log A. \qquad (8)$$

Naturally, the mth step will be taken only if, for $i = 1, 2, \ldots, m - 1$, the relations

$$\log B < k_i \log \frac{p_1}{p_0} + (i - k_i) \log \frac{1 - p_1}{1 - p_0} < \log A$$

have occurred.

For practical purposes the following simple method can be applied. Using the notation

$$c_1 = \frac{\log B}{\log \dfrac{p_1}{p_0} \dfrac{1 - p_0}{1 - p_1}}, \qquad c_2 = \frac{\log A}{\log \dfrac{p_1}{p_0} \dfrac{1 - p_0}{1 - p_1}},$$

$$d = \frac{\log \dfrac{1 - p_0}{1 - p_1}}{\log \dfrac{p_1}{p_0} \dfrac{1 - p_0}{1 - p_1}},$$

relation (8) takes the form

$$c_1 + md < k_m < c_2 + md. \tag{9}$$

Taking into account that $B < 1 < A$, which means that $\log B < 0 < \log A$, it follows that $c_1 < 0$ and $c_2 > 0$, and furthermore, obviously $d > 0$. The parallel straight lines $y_1 = c_1 + dx$ and $y_2 = c_2 + dx$, therefore enclose the origin $(0, 0)$ of the coordinate system. Those points (m, k_m), for which (9) is satisfied, i.e., for which the sequential procedure has to be carried out, are the (integer) lattice points situated between these two parallel lines in the first quadrant. The lattice points above the line of $y_2 = c_2 + dx$ are the rejection points (m, k_m) for which $k_m > c_2 + md$ holds; similarly the lattice points (m, k_m) below the straight line $y_1 = c_1 + dx$ correspond to the acceptance points, for which $k_m \leqslant c_1 + dm$. So it is easy to construct a procedure which shows what to do for each value $1, 2, \ldots$ of m.

The *operating characteristic* of the sequential test cannot be given explicitly in the binomial case, but in parametric form, as shown in Chapter III, Section 2.4.3, the OC function is of the form

$$L(p) \sim \frac{A^{h(p)} - 1}{A^{h(p)} - B^{h(p)}}.$$

The function $h(p)$ is defined by the relation

$$\sum_{(i)} \left(\frac{f(x_i; p_1)}{f(x_i; p_0)}\right)^{h(p)} f(x_i; p) = \left(\frac{f(0; p_1)}{f(0; p_0)}\right)^{h(p)} f(0; p) + \left(\frac{f(1; p_1)}{f(1; p_0)}\right)^{h(p)} f(1; p) = 1.$$

As $f(0; p_1) = p_1$, $f(0; p_0) = p_0$, we obtain

$$p\left(\frac{p_1}{p_0}\right)^{h(p)} + (1-p)\left(\frac{1-p_1}{1-p_0}\right)^{h(p)} = 1.$$

The functional form of $h(p)$ cannot be obtained explicitly from the equation; hence p will be expressed in terms of $h = h(p)$; i.e., the latter will be chosen as the independent parameter variable:

$$p(h) = \frac{1 - \left(\frac{1-p_1}{1-p_0}\right)^h}{\left(\frac{p_1}{p_0}\right)^h - \left(\frac{1-p_1}{1-p_0}\right)^h}.$$

Combining this with the relation

$$\bar{L}(h) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}$$

for $-\infty < h < \infty$, we obtain the OC curve in a $(p, \bar{L})$ coordinate system. $\bar{L}(.)$ indicates the approximate value of $L(.)$ obtained by replacing the values A and B by their approximate values $\frac{1-\beta}{\alpha}$ and $\frac{\beta}{1-\alpha}$ respectively. For any real value of h we obtain a point of the OC curve: for $h = +1$, the point $(p = p_0, L(p_0) = \bar{L}(1) = 1 - \alpha)$ is obtained, for $h = -1$, the point $(p = p_1, L(p_1) = \bar{L}(-1) = \beta)$.

The expected value of the sample number can be obtained from the formula given in the preceding Section. We know by now that in the binomial case

$$E(Z \mid p) = p \log \frac{p_1}{p_0} + (1-p) \log \frac{1-p_1}{1-p_0};$$

consequently

$$E(n \mid p) \sim \frac{L(p) \log \frac{\beta}{1-\alpha} + (1 - L(p)) \log \frac{1-\beta}{\alpha}}{p \log \frac{p_1}{p_0} + (1-p) \log \frac{1-p_1}{1-p_0}}.$$

For $p = p_0$ and $p = p_1$ this value can be determined exactly:

$$E(n \mid p_0) = \frac{(1-\alpha) \log \frac{\beta}{1-\alpha} + \alpha \log \frac{1-\beta}{\alpha}}{p_0 \log \frac{p_1}{p_0} + (1-p_0) \log \frac{1-p_1}{1-p_0}},$$

$$E(n \mid p_1) = \frac{\beta \log \dfrac{\beta}{1-\alpha} + (1-\beta) \log \dfrac{1-\beta}{\alpha}}{p_1 \log \dfrac{p_1}{p_0} + (1-p_1) \log \dfrac{1-p_1}{1-p_0}} .$$

Let us denote by N the sample size of the one-stage test for deciding between p_0 and p_1 at level $1 - \alpha$ and probability of error of the second kind β. The ratios

$$\frac{E(n \mid p_0)}{N} \quad \text{and} \quad \frac{E(n \mid p_1)}{N}$$

show the relative advantage in terms of average sample size of using a sequential plan instead of one-stage sampling. Unfortunately, these depend on p_0 and p_1 over α and β, and their tabulation is complicated. The corresponding values given for the Gaussian case are instructive for the binomial case as well. For normally and exponentially distributed variables this question will be considered for $\vartheta_0 < \vartheta < \vartheta_1$ as well; this is the disadvantageous case. If $\vartheta < \vartheta_0$ or $\vartheta > \vartheta_1$ the value of $E(n \mid \vartheta)$ will be smaller than for ϑ_0 and ϑ_1.

2.4.6 Sequential Test in the Case of a Normal Distribution

Let X be distributed according to the law $N(\vartheta; \sigma_0)$ with known variance σ_0^2. The simple null hypothesis

$$H_0\colon \vartheta = \vartheta_0$$

will be considered against the simple alternative

$$H_1\colon \vartheta = \vartheta_1$$

with the aid of the sequential procedure $(\vartheta_0 < \vartheta_1)$:

$$S = S\{\vartheta_0, \vartheta_1;\ A = A(\alpha, \beta);\ B = B(\alpha, \beta)\}.$$

For the density function of X_i we have

$$f(x; \vartheta_j) = \frac{1}{\sqrt{2\pi}\,\sigma_0}\, e^{-\frac{(x-\vartheta_j)^2}{2\sigma_0^2}} , \quad j = 0.1$$

from which we obtain.

$$z_i = \log \frac{f(x_i; \vartheta_i)}{f(x_i; \vartheta_0)} = \frac{\vartheta_1 - \vartheta_0}{\sigma_0^2}\, x_i + \frac{1}{2} \frac{\vartheta_0^2 - \vartheta_1^2}{\sigma_0^2} ,$$
$$i = 1, 2, \ldots .$$

The sum of the z_i's has the form

$$\sum_{i=1}^{m} z_i = \log \frac{p_{1m}}{p_{0m}} = \frac{\vartheta_1 - \vartheta_0}{\sigma_0^2} \sum_{i=1}^{m} z_i + \frac{m}{2} \cdot \frac{\vartheta_0^2 - \vartheta_1^2}{\sigma_0^2} .$$

Therefore, if the procedure has not come to an end in the first $m - 1$ steps, and we find at the mth step that

$$\frac{\vartheta_1 - \vartheta_0}{\sigma_0^2} \sum_{i=1}^{m} x_i + \frac{m}{2} \frac{\vartheta_0^2 - \vartheta_1^2}{\sigma_0^2} \geqslant \log A$$

has taken place, then H_0 will be rejected; while in the case

$$\frac{\vartheta_1 - \vartheta_0}{\sigma_0^2} \sum_{i=1}^{m} x_i + \frac{m}{2} \frac{\vartheta_0^2 - \vartheta_1^2}{\sigma_0^2} \leqslant \log B$$

H_0 will be accepted. The procedure will be continued by making an $(m + 1)$st observation if

$$\log B < \frac{\vartheta_1 - \vartheta_0}{\sigma_0^2} \sum_{i=1}^{m} x_i + \frac{\vartheta_0^2 - \vartheta_1^2}{\sigma_0^2} < \log A .$$

As σ_0, ϑ_0 and ϑ_1 are given, the procedure consists of comparing the sum $\sum_{i=1}^{m} x_i$ with the following upper and lower bounds:

$$\frac{2\,\sigma_0^2 \log A + (\vartheta_1^2 - \vartheta_0^2)\, m}{2(\vartheta_1 - \vartheta_0)}, \quad \frac{2\,\sigma_0^2 \log B + (\vartheta_1^2 - \vartheta_0^2)\, m}{2(\vartheta_1 - \vartheta_0)} .$$

As long as $\sum_{i=1}^{m} x_i$ lies between these two quantities, the procedure will be carried on; if $\sum_{i=1}^{m} x_i$ intersects one of them, the procedure terminates with a final decision.

In order to obtain the *operating characteristic function*, the value of $h(\vartheta)$ must be determined. Substituting the density function into the defining relation

$$\int_{-\infty}^{\infty} \left(\frac{f(x; \vartheta_1)}{f(x; \vartheta_0)} \right)^{h(\vartheta)} f(x; \vartheta)\, dx = 1 ,$$

the following equation will be obtained for the unknown $h(\vartheta)$:

$$\frac{1}{\sqrt{2\pi}\,\sigma_0} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_0^2} \{[\vartheta_1^2 - \vartheta_0^2 - 2x(\vartheta_1 - \vartheta_0)]h(\vartheta) + (x - \vartheta)^2\}}\, dx = 1 .$$

By carrying out the integration, we obtain the simple result

$$h(\vartheta) = \frac{\vartheta_0 + \vartheta_1 - 2\,\delta}{\vartheta_1 - \vartheta_0},$$

which does not depend on σ_0.

Using the approximate formula for the OC function given in Chapter III, Section 2.4.3 for the Gaussian case, we have

$$L(\vartheta) \sim \frac{\left(\dfrac{1-\beta}{\alpha}\right)^{\frac{\vartheta_0+\vartheta_1-2\vartheta}{\vartheta_1-\vartheta_0}} - 1}{\left(\dfrac{1-\beta}{\alpha}\right)^{\frac{\vartheta_0+\vartheta_1-2\vartheta}{\vartheta_1-\vartheta_0}} - \left(\dfrac{\beta}{1-\alpha}\right)^{\frac{\vartheta_0+\vartheta_1-2\vartheta}{\delta_1-\vartheta_0}}}.$$

For the *expected value of the sample size,* the determination of $E(Z \mid \vartheta)$ is needed. This can be obtained very easily and results in

$$E(Z \mid \vartheta) = \frac{1}{2}\,[2(\vartheta_1 - \vartheta_0)\vartheta + \vartheta_0^2 - \vartheta_1^2].$$

Substituting this into the approximate expression given in Chapter III, Section 2.4.4,we obtain

$$E(n \mid \vartheta) \sim \frac{L(\vartheta)\log A + (1 - L(\vartheta))\ \log B}{\dfrac{1}{2}\,[2(\vartheta_1 - \vartheta_0)\,\vartheta + \vartheta_0^2 - \vartheta_1^2]},$$

when $\vartheta \neq \dfrac{\vartheta_0 + \vartheta_1}{2}$.

Let us now denote by N the sample size of the one-sided u test, making a decision between H_0: $\vartheta = \vartheta_0$ and H_1: $\vartheta = \vartheta_1$ on the level $1 - \alpha$ and with probability of error of the second kind β. If $E(n \mid \vartheta)$ is the sequential test with the same error probabilities for the actual parameter value ϑ, then the ratios

$$\frac{E(n \mid \vartheta_0)}{N} \quad\text{and}\quad \frac{E(n \mid \vartheta_1)}{N}$$

may be considered as the *relative efficiency* of the sequential test with respect to the simple procedure; they do not depend on the values of ϑ_0 and ϑ_1. For certain pairs of α and β Table 28 gives the relative efficiency, which lies between 40–60% in important practical cases, showing the advantage of the sequential procedure in the case where ϑ_0 or ϑ_1 is true.

Table 28

THE PROPORTION OF THE AVERAGE SAMPLE NUMBER OF THE SEQUENTIAL PROBABILITY RATIO TEST, TO THE SAMPLE SIZE OF THE CORRESPONDING ONE-SIDED U TEST, IN MAKING A DECISION BETWEEN TWO VALUES OF THE EXPECTED VALUE OF A GAUSSIAN VARIABLE WITH KNOWN VARIANCE

α / β	H_0 *is true*			H_1 *is true*		
	0.01	0.03	0.05	0.01	0.03	0.05
0.01	0.42	0.49	0.53	0.42	0.39	0.37
0.03	0.39	0.46	0.50	0.49	0.46	0.45
0.05	0.37	0.45	0.49	0.53	0.50	0.49

The value of $\frac{E(n \mid \vartheta)}{N}$ will be higher when $\vartheta_0 < \vartheta < \vartheta_1$, and heuristic considerations lead to the conclusion that the highest value will be obtained in the middle point $\vartheta^* = \frac{\vartheta_0 + \vartheta_1}{2}$. Table 29 shows that even in this worst case the situation is not unacceptably disadvantageous. The table gives upper bounds for the relative efficiencies and at the same time absolute numbers for the expected sample sizes (see [B 80]).

Table 29

$\vartheta_1 = -\vartheta_0$	$\alpha = \beta$	$\frac{E(n \mid \vartheta^*)}{N}$ is not larger than	$E(n \mid \vartheta^*)$ is not larger than
1	0.01	1.34	7.2
	0.03	1.34	4.7
	0.05	1.38	3.8
0.5	0.01	1.13	23.9
	0.03	1.03	14.5
	0.05	1.00	10.8
0.25	0.01	1.03	89.1
	0.03	0.93	52.1
	0.05	0.88	38.0

2.4.7 Sequential Procedures for Testing Life Span in the Case of an Exponential Distribution

As already mentioned, the life span of equipments and systems can often be treated as if it were distributed according to the exponential law. For objects with a life span of this type, the following versions of the sequential method can be applied:

Let X be distributed according to the exponential law:

$$P(X \leqslant x) = 1 - e^{-\lambda x}, \; x \geqslant 0, \; \lambda > 0;$$

for the density function we have

$$f(x;\, \lambda) = \lambda e^{-\lambda x}$$

and the expected life length is

$$E(X \mid \lambda) = \frac{1}{\lambda}\,.$$

We will consider the null hypothesis

$$H_0\colon \lambda = \lambda_0,$$

against the alternative

$$H_1\colon \lambda = \lambda_1,$$

with $\frac{1}{\lambda_1} < \frac{1}{\lambda_0}$, i.e., $\lambda_1 > \lambda_0$.

According to the unique character of life testing, the single sample elements will not be examined one after the other. Instead, simultaneously n elements are examined over a period of time. The chosen value of n should be appropriately large, and the process starts at the same instant for all elements. In the course of the observation, the times of occurrences of failures must be registered as well as the total duration t of the procedure.

The following two versions of the testing procedure will be considered:

(a) each sample element is replaced after failing by a new one;

(b) no new sample element is added to the n pieces.

Here, when the experiment is to be observed continuously in time, the theory of Wald must be modified to deal with the case of continuous parameters. For this aspect we refer to the work by Epstein and Sobel [B 24], which is followed in our next treatment.

Two kinds of statistics will be introduced: the *total time on test* in the interval t on the one hand, and the total number r of the pieces which have failed in t on the other.

The total time on test will be in the case (a) $V(t) = nt$; in the case (b)

$$V(t) = \sum_{i=1}^{r} (n - i + 1)(x_i - x_{i-1}) + (n - r)\, x_r\, (t - x_r) =$$

$$= \sum_{i=1}^{r} x_i + (n - r)\, t,$$

where x_i denotes the life span of the sample element which is the ith to fail, and $x_0 = 0$.

The sequential procedure is based also on the probability ratio test and has the following form: The procedure will be continued as long as

$$\log B < r \log \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)\, V(t) < \log A$$

holds; H_0 will be accepted if, for the first time,

$$r \log \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)\, V(t) \leqslant \log B$$

is found; H_0 will be rejected if, for the first time,

$$\log A \leqslant r \log \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)\, V(t)$$

is obtained. As usual, $A = \dfrac{1 - \beta}{\alpha}$, $B = \dfrac{\beta}{1 - \alpha}$ can be used with error probabilities α and β.

In the case of (a) the sequential procedure may continue indefinitely; the expected duration of the test will be given below.

In the case of (b) it may happen that until time $t = x_n$, i.e., until the failure of the last sample item, no final decision between H_0 and H_1 will take place. If a sufficiently large value of n is chosen, the probability of this event can be made small, if one has a fair knowledge of the actual value of α. If, nevertheless, no final decision has been made, then the rejection of H_0 can be chosen. Applying this method, in the case of (b) the sequential procedure is used as a truncated test.

Let us introduce the following notation:

$$h_0 = \frac{-\log B}{\lambda_1 - \lambda_0}, \quad h_1 = \frac{\log A}{\lambda_1 - \lambda_0}, \quad s = \frac{\log \dfrac{\lambda_1}{\lambda_0}}{\lambda_1 - \lambda_0}.$$

Using this notation at time t^* the continuation of the procedure takes place, if

$$-h_1 + rs < V(t^*) < h_0 + rs.$$

The OC function can be written in the following parametric form:

$$\lambda(h) = \frac{h(\lambda_1 - \lambda_0)}{\left(\frac{\lambda_1}{\lambda_0}\right)^h - 1}, \qquad L(h) = \frac{A^h - 1}{A^h - B^h}.$$

The value of L is obtained immediately for five different values of λ:

$$\begin{array}{llllll} \frac{1}{\lambda} = & 0 & \frac{1}{\lambda_1} & s & \frac{1}{\lambda_0} & \infty \\ L(\lambda) = & 0 & \beta & \frac{\log A}{\log A - \log B} & 1-\alpha & 1 \end{array}$$

Instead of the expected sample size, in this case the observed total time on test and the total number of failed pieces will be considered. It can be shown that in both of the cases (a) and (b),

$$E(V(t) \mid \lambda) = \frac{1}{\lambda} E(r \mid \lambda)$$

is valid. As a good approximation we may use

$$E(r \mid \lambda) \sim \begin{cases} \dfrac{h_1(1 - L(\lambda)) - h_0 L(\lambda)}{s - \dfrac{1}{\lambda}}, & \text{when } \dfrac{1}{\lambda} \neq s, \\ \dfrac{h_0 h_1}{s}, & \text{when } \dfrac{1}{\lambda} = s. \end{cases}$$

The value of $E(r \mid \lambda)$ can be given in a simple form for λ_0, λ_1 and for $\lambda = \frac{1}{s}$. The latter yields information about the test procedure in the worst case. Table 30 gives the value of $E(r \mid \lambda)$ for different values of α, β and $\frac{\lambda_1}{\lambda_0} = k > 1$.

Finally, we remark that in the case of (a) the expected length of time until the termination of the procedure can be given very easily, as $V(t) = nt$:

$$E(t \mid \lambda) = \frac{1}{n\lambda} E(r \mid \lambda).$$

Example: The problem concerns the expected life time of bulbs: $H_0 = \frac{1}{\lambda_0} = 7\,500$ hours will be considered against $H_1 = \frac{1}{\lambda} = 2\,500$ hours at level $1 - \alpha = 0.95$ and with $\beta = 0.05$. The testing is performed on $n = 100$ bulbs simultaneously; the failed items are replaced by new ones.

TABLE 30

THE EXPECTED NUMBER OF FAILED ITEMS AT TERMINATION IN THE CASE OF THE DESCRIBED SEQUENTIAL PROCEDURE FOR EXPONENTIAL LIFE SPAN

$k = \frac{\lambda_1}{\lambda_0}$		3/2		2		5/2		3	
λ	β \ α	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
λ_1	0.01	62.4	40.3	23.3	15.1	14.2	9.2	10.4	6.7
	0.05	60.4	36.7	22.6	13.7	13.8	8.4	10.1	6.1
$\frac{1}{s}$	0.01	128.0	82.7	43.9	28.3	25.1	16.2	17.5	11.3
	0.05	82.7	52.7	28.3	18.0	16.2	10.3	11.3	7.2
	0.01	47.6	44.2	14.7	13.6	7.7	7.66	5.0	4.6
λ_0	0.05	30.8	28.0	9.5	8.6	5.0	4.54	3.2	2.9

As

$$k = \frac{\lambda_1}{\lambda_0} = 3, \quad \lambda_1 - \lambda_0 = \frac{1}{3750},$$

$$A = \frac{0.95}{0.05} = 19 \quad B = \frac{0.05}{0.95} = \frac{1}{19},$$

consequently, the following relation is the deciding one with respect to the test procedure:

$$\log \frac{1}{19} = -2.9452 < 1.0985\, r - \frac{V(t)}{3750} < \log 19 = 2.9452.$$

The expected length of time of the test procedure can be obtained in the following way: as $\log B = -\log A$,

$$h_0 = h_1 = \frac{\log A}{\lambda_1 - \lambda_0} = 3750 \log 19 \sim 12045$$

$$s = 3750 \log 3 \sim 4115;$$

hence, we obtain

$$E(t|\,\lambda_0) = \frac{1}{\lambda_0\, n} E(r\,|\,\lambda_0) \sim \frac{h_0(2\alpha - 1)}{s - \frac{1}{\lambda_0}} \frac{1}{n\,\lambda_0} \sim 240 \text{ hours}$$

$$E\left(t\,\middle|\,\frac{1}{s}\right) = \frac{s}{n} E(r\,|\,s) \sim \frac{h_0}{s^2} \frac{s}{n} \qquad \sim 350 \text{ hours}$$

$$E(t|\lambda_1) = \frac{1}{\lambda_1 n} E(r|\lambda_1) \sim \frac{h_0(2\alpha - 1)}{\dfrac{1}{\lambda_1} - s} \frac{1}{n\lambda_1} \sim 167 \text{ hours.}$$

By using more accurate methods, these values can be improved, but even here it can be observed that the expected duration of the sequential test procedure is only a small fraction of the expected life span of the bulbs under consideration.

2.4.8 Truncation and Grouping in the Sequential Test Procedure

Two remarks will be made here concerning the practical performance of the sequential probability ratio test.

The sequential procedure is usually not carried on for an unrestrictedly indefinitely long time; instead, the test is truncated. For this purpose an integer n_0 is chosen sufficiently large that the process will come to an end with high probability before reaching the sample number n_0:

$$P(n < n_0 \mid H_i) \geqslant 1 - \varepsilon, \; i = 1, 2,$$

(e.g., $\varepsilon = 0.05$ or $\varepsilon = 0.01$). This means that with a probability higher than $1 - \varepsilon$ the procedure leads to a final decision in less than n_0 steps, whichever of the hypotheses is true. If we wish, however, to definitely finish the procedure after the n_0th step, we have to determine how this is to be done when at the n_0th step the sequential stipulation does not lead to the termination of the test. This may be carried out in the following way: let us accept H_0 if $\sum_{i=1}^{n_0} z_i \leqslant 0$, and let us reject H_0 if $\sum_{i=1}^{n_0} z_i > 0$.

TABLE 31
UPPER BOUNDS FOR THE ERROR PROBABILITIES
IN A SEQUENTIAL PROCEDURE

n_0	$\alpha = 0.01$	$\beta = 0.01$	$\alpha = 0.01$	$\beta = 0.05$	$\alpha = 0.05$	$\beta = 0.05$
1000	0.020	0.020	0.033	0.070	0.095	0.095
1500	0.0125	0.0125	0.0175	0.0565	0.069	0.069
2000	0.0100	0.0100	0.0120	0.0520	0.058	0.058
2500	0.0100	0.0100	0.0110	0.0510	0.054	0.054
3000	0.0100	0.0100	0.0100	0.0500	0.052	0.052

The truncation of the sequential procedure obviously modifies, then, the error probabilities α and β. Let us denote these modified values by $\alpha(n_0)$ and $\beta(n_0)$. In the case of Gaussian variables Table 31 shows upper bounds of $\alpha(n_0)$ and $\beta(n_0)$ for different combinations of n_0 and the original error probabilities α and β, all for a sequential procedure with average sample size 1000. As can be seen, the error probabilities decrease significantly from 1000 to 2000, and show for $n_0 = 2000$, in most of the practical cases, no significant deviation from the original values.

A second modification of the sequential test procedure is *the grouping*; the sample elements are examined, not one by one, but in groups of a given size, say v. This means that first the probability ratio $\frac{p_{1v}}{p_{0v}}$ will be considered; finding

$$B < \frac{p_{1v}}{p_{0v}} < A ,$$

the next v sample elements will be examined.

It can be easily seen that by using grouping, an increase will take place in the expected value of the sample size; but this increase is not necessarily large if v is moderate. As, however, the examination of v single sample element takes more time than the examination of a group, we may regain the time needed for the increased sample number. We do not go into details in this respect, but wish to mention only that the grouping affects the error probabilities to a very small extent only. We refer once again to Wald [A 48].

2.5 ACCEPTANCE PLANS AND THE PRODUCTION AS A STOCHASTIC PROCESS

2.5.1 Stochastic Relation Between Sample and Lot

We begin by mentioning a remarkable example, due to Kolmogorov, concerning the stochastic relation between the number of defective items in a sample and the number of defective items in the non-examined part of the lot. This will demonstrate that from the point of view of an acceptance plan it is of importance whether these quantities (as random variables) are correlated or not. Suppose we have a sequence of lots containing 100 pieces each. The acceptance procedure consists of the examination of 10 items taken randomly from the lot; if any defective item is found in the sample, the lot is to be rejected; otherwise it is to be accepted. Suppose now that each lot contains a single defective piece. In this case all of our inferences will be incorrect: we accept the non-examined 90 items if no defective item is contained in the sample and we reject them if all of the sample pieces are good.

Although the assumptions of this example are rather artificial, nevertheless, they call our attention to the problem.

Our earlier considerations have shown that in choosing a sampling plan we take into account (if possible) our information on the average quality level of the production process. Let us now see how the sampling plan will be influenced by our assumptions concerning the production process.

Suppose that the goods are produced by automatic equipment and the consecutive pieces turn out to be defective with probability p and good with probability $1 - p$ *independently*. In this case a lot of size N can be considered as a random sample from an infinite population. More precisely, a random sample of size n taken from the lot and the remainder of the lot of size $N - n$ can be considered as two independent random samples from the same infinite population containing defective items with probability p. Consequently, the number of defective items in the sample and in the remainder are two independent random variables; *no inference at all can be made from the one to the other*. Despite this surprising statement, there is no error in our sampling procedure: considering the sample of size n, we may draw statistical inference concerning p. If p is accepted as the true value, we know that in the remainder the number of defective items follows the binomial distribution with parameters $(N - n)$ and p. Consequently, we are in possession of good information—at least on the average—about the accepted number of defective pieces.

As in our earlier example the dependent case has a different feature, we can immediately draw inference about the remainder of the lot in a positive or negative sense, according to whether the mentioned two random variables are positively or negatively correlated. The above example can be modified to a more realistic situation. Suppose that a firm produces a certain instrument and the product turns out to be excellent, without any defective items. But the firm has a large number of bad pieces from earlier production of the same product and—let us hope that such a firm does not exist, but only for the sake of an illustration we assume for a moment that—it wishes to sell these as well. If it mixes s bad products into each delivered lot, then—using a simple sampling plan—for the values $\varkappa_n$ and $\varkappa_{N-n}$, i.e., for the number of bad items in the sample and in the remainder, the relation $\varkappa_n + \varkappa_{N-n} = s$ holds; consequently, the better the sample, the worse the remainder part will be.

We now call attention to the following circumstance: From a probabilistic point of view we may start with a fixed lot containing a fraction p defective, which is unknown and which is tested by means of a sampling procedure. In this case the underlying mass phenomenon is the random choice of the n sample elements from the (large) lot, in which the number of defective items can be considered distributed according to the binomial law. This concept can be applied, e.g., if we have to deal on a single occasion with one lot only, or if we are in the same position in a series of cases but we do not know anything at all about the variability of p.

On the other hand, taking into account a production process in which defective items occur according to a given stochastic pattern (dependent or independent), the lots themselves must be considered as samples from a large population with definite statistical properties. We aim now at showing how a certain knowledge of the stochastic process influences the sampling plan; this means that such considerations help us in constructing sampling plans better fitted to real situations. We will treat the Pólya process (see Chapter II, Section 4.6.6); the succeeding pieces will turn out to be good or defective during the production according to this law. We can motivate our selection—as we have done in the section just cited—in the following way: in many practical situations it lies on hand to suppose that if a piece turns out to be good then we have a larger probability for the next piece to be good; conversely, a defective item will be followed with higher probability by another defective item than by a good one.

We turn now to the examination of the corresponding acceptance plans. These investigations will lead to a further important consequence: they show *that the OC function is not always the appropriate tool in making judgements concerning sampling plans.*

2.5.2 Sample and Lot in the Case of a Pólya Sequence

Let us consider a production process where the pieces turn out successively to be good or defective with probability $1 - p$ and p, respectively ($0 < p < 1$), according to the Pólya sequence. This means that if ξ_i is the random variable for which

$$\xi_i = \begin{cases} 1 \text{ if the } i\text{th piece is defective} \\ 0 \text{ if the } i\text{th piece is good,} \end{cases}$$

then

$$P(\xi_{n+1} = 1 \mid s_n = \xi_1 + \xi_2 + \ldots + \xi_n = s) = \frac{p + sq}{1 + nq} \,,$$

where for the parameter q the condition $q > 0$ will be assumed. The case $q = 0$ corresponds to the independent case, i.e., to the binomial sequence.

Consider now a lot of size N from which a sample of size $n(< N)$ will be taken. First of all, let us make a remark concerning the choice of the sample. The Pólya sequence forms a system of so-called equivalent events, which means that the probability

$$P(\xi_{i_1} = \varepsilon_1,\ \xi_{i_2} = \varepsilon_2, \ldots, \xi_{i_n} = \varepsilon_n),\ \varepsilon_j = 0 \text{ or } 1,\ j = 1, 2, \ldots, n,$$

does not depend on the order of the indices in the permutation $(i_1, i_2, \ldots, i_n)$. Consequently it is quite unimportant whether the sample is chosen randomly or the first n or the last n, etc. pieces are taken. (The choice is made among the indices and not among the pieces! Thus it cannot be influenced by any product being good or defective.) The number of defective items $s_n = \xi_{i_1} + \xi_{i_2} + \ldots + \xi_{i_n}$ in the sample will always have the distribution

$$P(s_n = s) = \binom{n}{s} \frac{\prod\limits_{i=0}^{s-1} (p + iq) \prod\limits_{j=0}^{n-s-1} (1 - p + jq)}{\prod\limits_{l=0}^{n-1} (1 + lq)} \,, \tag{10}$$

$$s = 0, 1, 2, \ldots n.$$

For the expected value and variance we have

$$E(s_n) = np, \quad D^2(s_n) = np\,(1 - p)\,\frac{1 + nq}{1 + q} > np(1 - p).$$

We may determine the distribution of the number of defectives in the remaining part of the lot under the condition that the sample contains s defective pieces. This probability is, according to the theorem of Bayes,

$$P(s_{N-n} = x \mid s_n = s) = \frac{P(s_n = s \mid s_{N-n} = x)\, P\,(s_{N-n} = x)}{\sum\limits_{i=0}^{N-n} P(s_n = s \mid s_{N-n} = i)\, P(s_{N-n} = i)}$$

$$x = 0, 1, 2, \ldots, n.$$

Having a lot with $s_N = s_{N-n} + s_n = x + s$ defective items, the probability that a random sample of size n taken from it contains s defective items is—according to the hypergeometric distribution—

$$P(s_n = s \mid s_N = x + s) = \frac{\binom{x+s}{s}\binom{N-s-x}{n-s}}{\binom{N}{s}} .$$

Using this formula, a simple transformation leads to the result

$$P(s_{N-n} = x \mid s_n = s) = \binom{N-n}{x} \frac{\prod_{i=0}^{x-1} (p_1 + i q_1) \prod_{j=0}^{N-n-x-1} (1 - p_1 + j q_1)}{\prod_{l=0}^{N-n-1} (1 + l q_1)}$$

$$x = 0, 1, 2, \ldots, N - n,$$

where

$$p_1 = \frac{p + sq}{1 + nq}, \quad q_1 = \frac{q}{1 + nq} .$$

This result has the interpretation that the *posterior* distribution of the number of the defectives in the lot is also a Pólya distribution, but now with parameters depending on s. The conditional expected value of s_{N-n} under the condition that $s_n = s$ is

$$E(s_{N-n} \mid s_n = s) = (N - n)\, p_1 = (N - n) \frac{p + sq}{1 + nq} .$$

The correlation coefficient can be determined in an elementary way as well, and has the form

$$\operatorname{corr}\ (s_{N-n}, s_n) = \frac{q}{\sqrt{\left(q + \frac{1}{n}\right)\left(q + \frac{1}{N-n}\right)}} .$$

This means that the number of defective items in the sample and in the remaining part of the lot are, in the case of a Pólya process with $q > 0$, positively correlated: more defectives in the sample implies, on the average, more defectives in the remaining part of the lot.

2.5.3 On the Acceptance Plans in the Case of a Pólya Sequence

The effect of the stochastic character of production on the acceptance plans will be considered in the case of a Pólya sequence. We discuss the problem in two steps: first the $(n, 0)$ plans will be treated, they show a definite trend in the deviation of the sample size from the independent case. Concerning $(n, 0)$ plans with $c \geqslant 1$, the situation is not so unique; it depends on the parameters.

(a) Suppose that lots are to be examined and we choose a sampling plan with zero acceptance number in the sample, i.e., an $(n, 0)$ plan. Suppose that for lots with fraction defective $p = p_0$, the prescribed probability of the error of the first kind makes $1 - \alpha = 0.95 = 95\%$. From this the size n of the sample is determined in the following way: We assume that the second parameter q of the Pólya sequence is given. In this case the OC function of the plan has the form

$$L(p; q, n, 0) = P(s_n = 0) =$$

$$\frac{\prod_{j=0}^{n-1} (1 - p + jq)}{\prod_{l=0}^{n-1} (1 + lq)} = (1 - p)\left(1 - \frac{p}{1+q}\right) \cdots \left(1 - \frac{p}{(1 + (n-1)q)}\right) .$$

The OC function of the corresponding plan in the independent case is given by the binomial distribution

$$L(p;\ 0,\ n,\ 0) = (1 - p)^n .$$

Since $p > \dfrac{p}{1 + iq}$ holds for $i = 1, 2, \ldots, n - 1$, it follows that for each value of p and for $q > 0$,

$$L(p;\ 0, n, 0) > L(p;\ q, n, 0)$$

is valid. According to our requirements, the relation

$$L(p_0;\ q, n, 0) = 1 - \alpha = 0.95$$

must be satisfied. This relation determines n. On the other hand—using the binomial model as above—

$$0.95 = L(p_0;\ 0, n', 0)$$

must be satisfied. As

$$L(p_0;\ 0, n', 0) < L(p_0;\ q,\ n', 0)$$

is valid, in the case of equality the relation must hold. Thus we have achieved the interesting and important result that *an $(n, 0)$ plan with given probability for the error of the first kind must have a larger sample size for the case of a Pólya sequence than for the corresponding binomial sequence.*

Now the general case will be considered.

(b) We begin with the extension of the above statement to an arbitrary (n, c) plan $(c \geqslant 1)$: the same assertion is true if q is sufficiently small $(q < q_0(n, c))$ as in the case of (a). This means that to ensure the same probability for the error of the first kind, a larger sample size is needed for $q > 0$ than for the independent case. The proof of this theorem was given by E. Vas (unpublished).

The general case is not as simple and easy to survey. According to a theorem of Tusnády [B 64] the following statement is true. Suppose that for p_1 and $p_2 (> p_1)$ and for some $q (\geqslant 0)$, there exists a sampling plan (n, c) with the property

$$L(p_1;\ q, n, c) = 1 - \alpha,\ L(p_2;\ q, n, c) = \beta.$$

Further let $q' > q$. Then for any acceptance plan (n', c') which is better than this in the sense that

$$L(p_1;\ q', n', c') \geqslant 1 - \alpha,\ L(p_2;\ q', n', c') \leqslant \beta,$$

the sample size n' must be larger than n.

On the other hand, if q is large, i.e., the stochastic connection between the defective items in the sample and in the lot is very strong, then the sample size can be smaller. For this reason we turn to the extreme case $q = \infty$. Since

$$P(\xi_2 = 1) = \frac{p + s_1 q}{1 + q} \rightarrow s_1, \quad \text{as} \quad q \rightarrow \infty\,,$$

the probability for the second piece to be defective is one in the case where the first piece has been defective, etc. In this case a lot will contain either good items or bad items only. This means that for getting full information on the quality of the lot the sample size $n = 1$ is sufficient. If the single sample element is good, we accept the lot; otherwise we reject it with probability 1.

In this case we have for the OC function

$$L(p;\ \infty, 1, 0) = 1 - p,$$

i.e., we accept the lot with probability $1 - p$ when p is the true fraction defective. Does this involve a contradiction? By no means. Here the parameter p of the production process means that the production leads, with probability p, to purely bad lots and, with probability $1 - p$, to purely good lots. Thus, in the long run, we indeed accept the lots with probability $1 - p$. The *conditional* probability of acceptance and rejection turns out to be one at each step, but *not on the average.*

This consideration shows that in the Pólya case the OC function is not an appropriate tool for choosing a sampling plan. The information contained in the sample refers, in many cases, primarily to the fraction defective in the actual lot and not so much the process fraction defective p. In the last case ($q = \infty$) we are fully informed on the whole lot from a single sample element, but we need a sequence of lots to estimate p.

In our considerations above we were only able to show that the stochastic properties of the production process must be taken into account in the construction of appropriate acceptance plans. Positive results are given here for very special cases only, but these also show the importance of the problem and—at least we hope so—call attention to the need for further theoretical investigations.

3 RELIABILITY THEORY

In the last decades a characteristic feature of industrial development was the tendency towards an increasing complexity of equipment, coupled with strict working conditions. This resulted in higher requirements for the reliability of some products. This circumstance initiated the development of the theory of reliability.

In the applications of probability theory and mathematical statistics the problems concerning the behavior of complex systems or equipment are of special importance.

The failure of the components is of random character in the field and in preliminary tests as well. It is evident, therefore, that probability theory and mathematical statistics are of central importance in the theory of reliability.

The term "reliability"—if used in a quantitative sense—means the probability of performing a function for a given period of time under certain well-defined conditions.

Problems of reliability theory may be classified according to two major groups: they may concern the satisfactory product performance over a period of time or the probability that a product or device will work without failure under given conditions when called upon to do so. In the first group of problems the time to failure (life length) and its probability distribution are of basic importance.

3.1 FAILURE DISTRIBUTION

Having a product, let us consider the random interval beginning with the moment the product starts to work and ending at the moment of its (first) failure. (For the sake of simplicity, let us assume that failure may be the only cause of the termination of work.) This positive random variable has a basic role in reliability investigations. Its name is *time to* (*first*) *failure*. After renewal or replacement has occurred it is called the *time between failures* (TBF); for nonreparable products also the terms *life length* or *life time* are used. Their expectations are called the *mean time to* (*first*) *failure* (MTTF), *mean time between failures* (MTBF) and *average life time*, respectively.

In most cases this variable may be described by the distribution function $F(t)$, or the density function $f(t)$ may be used for the characterization of this distribution. In reliability investigations, however, the use of a third function is often more suitable: this is *the hazard function* (*hazard rate*, *instantaneous failure rate*):

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{11}$$

defined for values of t for which $F(t) < 1$.

Hence, $\lambda(t)$ is the conditional density of the failure probability at the instant t, given the condition that the product has not failed prior to this time.

If $\lambda(t)$ is given, $F(t)$ can be expressed in terms of it. We obtain

$$\int_0^t \lambda(x)\,dx = \int_0^t \frac{f(x)\,dx}{1 - F(x)} = \int_0^{F(t)} \frac{dx}{1 - x} = -\ \log\,[1 - F(t)]$$

rom (11) by integration, i.e.

$$F(t) = 1 - \exp\left[-\int_0^t \lambda(x)\,dx\right].$$

The function $R(t) = 1 - F(t)$ is often called *reliability function*.

Experience shows that the shape of the curve represented in Fig. 54 is typical of the hazard function of many products. Practically speaking, it consists of three parts; in the first part the hazard function is decreasing, in the second it is approximately constant and in the third it is increasing. The first period is that of the early failures. Often the products are "burnt in" prior to actual use in order to eliminate this period.

Inhomogeneities in the material may increase the length of this time interval.

The second part is normally regarded as the period of useful life. Mostly "chance failures" occur in this period.

The third period is dominated by the "wearout failures".

Sometimes the shape of the curve may be simpler than that of Fig. 54. In other cases we may approximate the failure distribution by some simpler formula, aware of the fact that the approximation is good only for a limited period.

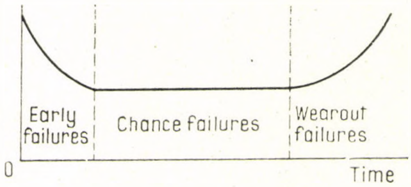


FIG. 54. Typical hazard rate curve

Approximation by the exponential distribution is of special importance in this field. This distribution is characterized by a constant hazard rate, this being the reciprocal value of the expectation. This means that if the product has survived to time t, then the probability of survival for a next increment of time is the same as if just been placed into service (i.e., products having an exponentially distributed failure distribution "do not grow old"). This exact situation very rarely occurs in practice. This model, however, may be a useful approximation.

Experience shows that the failure distributions of equipment or systems having a large number of components can be well approximated by exponential distributions. The reason for this lies in the fact that the time between system failures is small compared to the times-to-failure of the components, and the aging of the components is not considerable within this short period.

Another distribution often used for approximating failure distributions is the Weibull distribution (Chapter I, Section 2.9.7), characterized by a hazard rate of the form γx^{β}. If β is positive (negative) the hazard rate is increasing (decreasing). The special case $\beta = 0$ yields the exponential distribution.

The exponential, Gamma and Weibull distributions may be generalized by introducing a location parameter: the "guarantee" time. No unit can fail before this time: the distribution is shifted by this quantity. For example, the density function of the shifted exponential distribution is:

$$f(x) = \begin{cases} 0 \quad , & \text{if } x < a , \\ \lambda e^{-\lambda(x-a)}, & \text{if } x \geqslant a , \end{cases}$$

where a is the "guarantee" time (location parameter) and λ is the hazard rate.

In addition to these two types (i.e., exponential and Weibull distributions), normal, lognormal, Gamma distributions are also used for approximating certain failure distributions. Figures 55–59 represent the distribution function, density function and hazard rate function of distributions belonging to these five types.

Frequently, units that fail on test are replaced by new ones of similar kind. In such circumstances we may ask how many failures have occurred in the interval $(0, t)$.

Let us denote the nth convolution of the failure distribution $F(t)$ by $F_n(x)$, i.e., $F_1(x) = F(x)$,

$$F_n(x) = \int_0^x F_{n-1}(x - y)dF(y), \quad n = 2, 3, \ldots$$

Then $F_n(t)$ is the probability that at least n failures happened during the interval $(0, t)$; i.e., the probability that ν_t, the number of failures in this interval, is exactly n, is

$$P(\nu_t = n) = F_n(t) - F_{n-1}(t), \quad n = 1, 2, \ldots$$

where $F_0(t) = 0$.

Therefore

$$H(t) = E(\nu_t) = \sum_{n=1}^{\infty} n[F_n(t) - F_{n-1}(t)] = \sum_{n=1}^{\infty} F_n(t).$$

$H(t)$ is called the *renewal function*. Its derivative:

$$h(t) = H'(t) = \sum_{n=1}^{\infty} f_n(t),$$

where $f_n(t) = F'_n(t)$, is the *renewal density*. For small values of Δt, the probability that a failure takes place in the interval $(t, t + \Delta t)$ is $h(t)\Delta t$.

In the case of an exponential distribution, $h(t) = \lambda$, i.e., the renewal density is constant and identical with the failure rate. The formulae for $H(t)$ and $h(t)$ in the case of normal and Gamma distributions are given in [A 19], pp. 100—101.

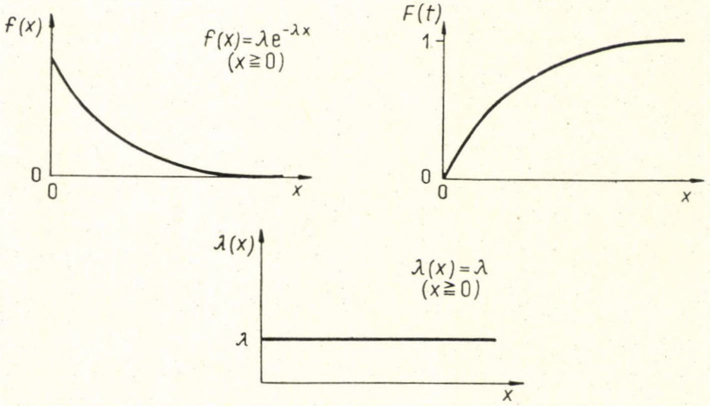


FIG. 55. Distribution function, density function and hazard rate function for the exponential distribution

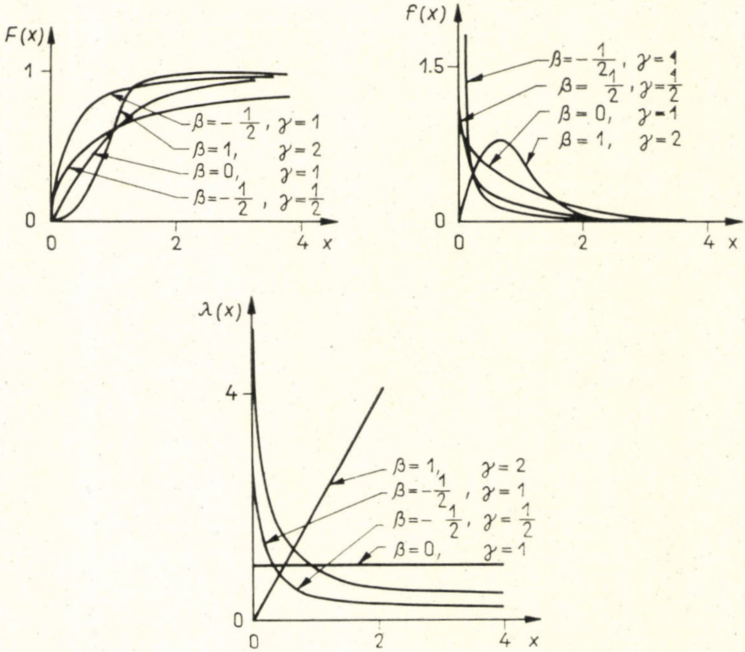


FIG. 56. Distribution function, density function and hazard rate function for the Weibull distribution

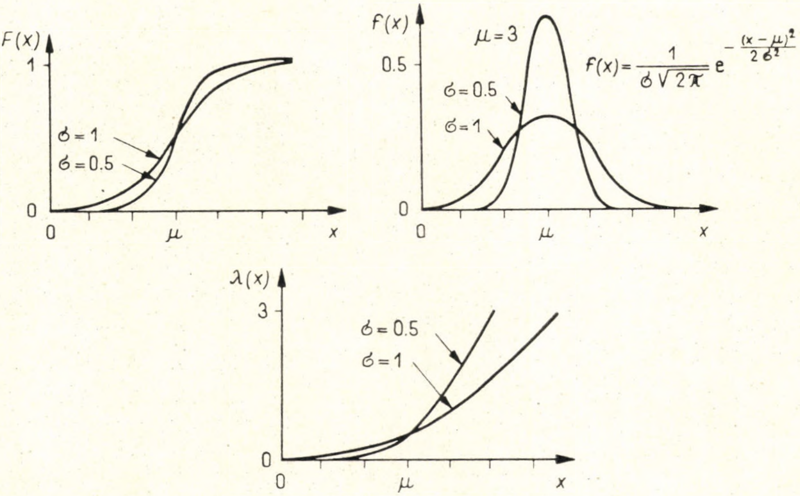


FIG. 57. Distribution function, density function and hazard rate function for the normal distribution

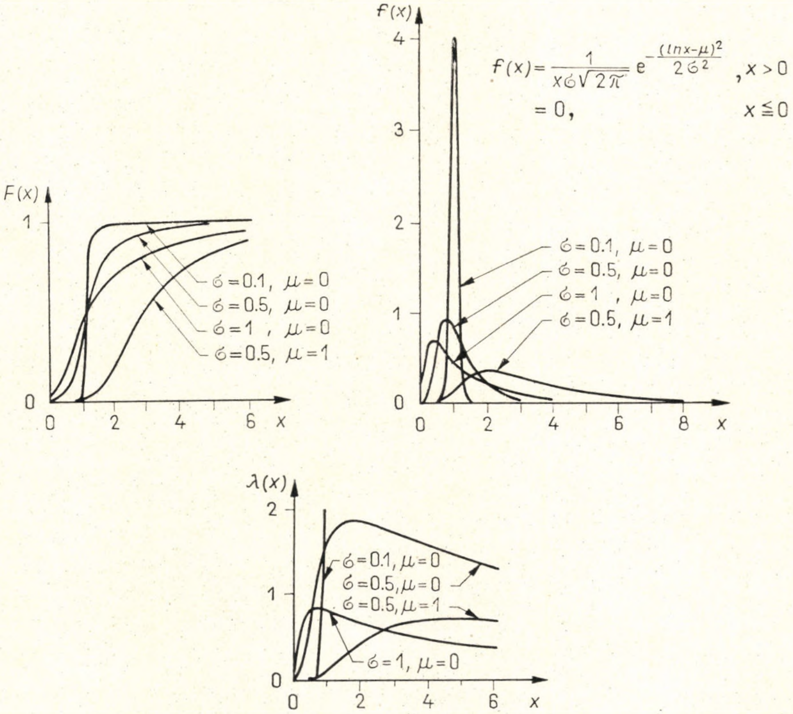


FIG. 58. Distribution function, density function and hazard rate function for the log-normal distribution

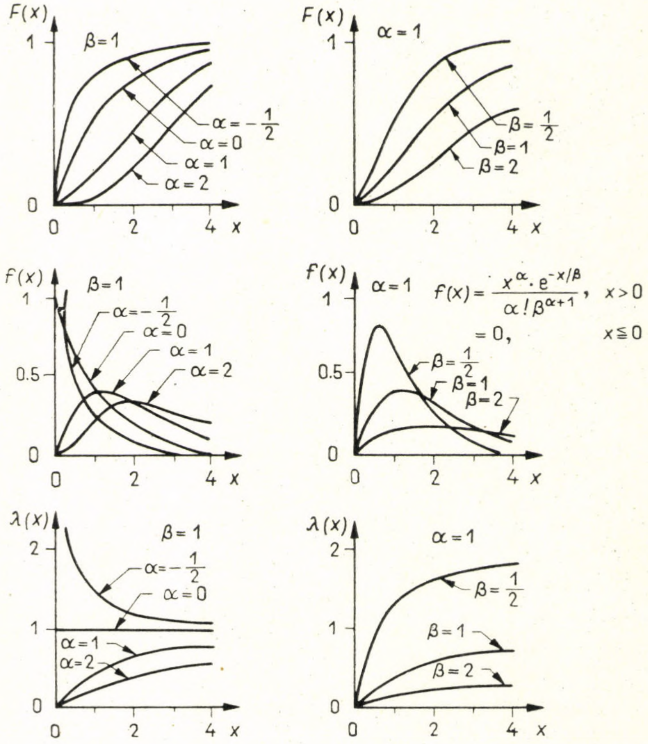


FIG. 59. Distribution function, density function and hazard rate function for the gamma distribution

For any distribution

$$\lim_{t\to\infty} \frac{H(t)}{t} = \frac{1}{\tau_0}$$

holds and under rather mild conditions

$$\lim_{t\to\infty} h(t) = \frac{1}{\tau_0}$$

where τ_0 is the expectation of the failure distribution (see, e.g., [A 19], pp. 102–103).

In some problems the failure distribution is *discrete* (e.g., the number of cycles in fatigue studies). As approximations, negative binomial, Poisson and other distributions are used.

3.2 STATISTICAL EVALUATION OF EMPIRICAL FAILURE DATA

The analysis of data for estimating unknown parameters or performing statistical tests in reliability applications can be accomplished according to the general rules. Some specific methods are, however, most frequently used in reliability problems. These are the specific methods of the statistical estimation and testing of the hazard rate and reliability, problems connected with the exponential and Weibull distributions. In life testing we often wish to evaluate the data obtained before all units investigated have failed. In this case we have, in the terminology of statistics, a truncated sample.

Below, we briefly discuss a few important problems.

(a) *Exponential distribution, complete sample.* For the hazard rate λ we usually use the unbiased estimator ([A 19], p. 177)

$$\hat{\lambda} = \frac{n-1}{\sum_{i=1}^{n} t_i}$$

where n is the sample size, i.e., the number of all tested units and $t_1, t_2, \ldots, t_n$ are the observed lifetimes. It agrees asymptotically with the maximum-likelihood estimator of the hazard rate λ.

$\sum_{i=1}^{n} t_i/\lambda$ is the sum of n independent exponentially distributed variables with unit parameter. Therefore, it is Gamma-distributed (see Chapter II, Section 2.9.5). Using this fact, we can prove that $\hat{\lambda}$ is unbiased, with variance

$$D^2(\hat{\lambda}) = \frac{\lambda^2}{n-2}$$

and we can construct confidence intervals and statistical tests for λ. The tables of the Gamma distribution [C 9] being less available, χ^2 tables are usually used for this purpose. Alternatively, Table 7 of [A 19] can be used. With χ^2 tables, the confidence interval is yielded by

$$P\left(\frac{\chi^2_{\varepsilon/2,\,2n}}{2T} < \lambda < \frac{\chi^2_{1-\varepsilon/2,\,2n}}{2T}\right) = 1 - \varepsilon$$

where $\chi^2_{\delta,k}$ stands for the δ quantile of the χ^2 distribution with k degrees of freedom, $T = \sum_{i=1}^{n} t_i$. (Cf. II. 2.9.6.)

If one wishes to test the hypothesis H_0: $\lambda = \lambda_0$ against the alternative H_1: $\lambda \neq \lambda_0$, calculate $2\lambda_0 T$ and accept H_0 if $\chi^2_{\alpha/2, 2n} < 2\lambda_0 T < \chi^2_{1-\alpha/2, 2n}$; otherwise reject it. Here α is the size of the test.

The reliability $R(t) = e^{-\lambda t}$ may be estimated by $e^{-\hat{\lambda} t}$. This estimator is biased; an unbiased estimator is given in [B 62].

(b) *Exponential distribution, truncated sample.* Suppose n units are placed on test, and the test is terminated at a predetermined time t; the number of units failing prior to this instant is denoted by r, and the corresponding failure time data are $0 \leqslant t_1 \leqslant t_2 \leqslant \dots \leqslant t_r \leqslant t$. In this case the maximum-likelihood estimator for the hazard rate is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + (n - r)t}.$$

The denominator is the "total time on test" accumulated by all items. This estimator is asymptotically unbiased.

The distribution of $\hat{\lambda}$ is rather complicated. There is no unbiased estimator in this case. The basic difficulty arises by the lack of a one-variate sufficient statistic; the pair of quantities $T = \sum_{i=1}^{r} t_i + (n-r)t$ and r form a sufficient statistic. The variance of $\hat{\lambda}$ can be estimated by the formula

$$D^2(\hat{\lambda}) \sim \frac{r}{\left[\sum_{i=1}^{r} t_i + (n - r)t\right]^2}.$$

Epstein [B 22–23] has proposed two modifications of the testing procedure; by these the sufficient statistic becomes one-dimensional and its distribution has a simpler form.

(c) In the first of these modifications the investigation time t is not given in advance but the test is continued until $r = r_0$ failures occur (r_0 is given). In this case

$$\hat{\lambda} = \frac{r_0 - 1}{\sum_{i=1}^{r_0} t_i + (n - r_0)t_{r_0}}$$

is an unbiased estimator; its variance is

$$D^2(\hat{\lambda}) = \frac{\lambda^2}{r_0 - 2}$$

if $r_0 > 2$. Note that in this case $\frac{1}{\lambda}\left[\sum_{i=1}^{r_0} t_i + (n - r_0)\, t_{r_0}\right]$ is Gamma-distributed with shape parameter r_0 by which the above formula can be deduced. A confidence interval of confidence level $1 - \varepsilon$ is the following:

$$\frac{\chi^2_{\varepsilon/2,\, 2r_0}}{2T} < \lambda < \frac{\chi^2_{1-\varepsilon/2,\, 2r_0}}{2T}$$

where $T = \sum_{i=1}^{r_0} t_i + (n - r_0)t_{r_0}$.

(d) According to the second modification neither t nor r but $T = \sum_{i=1}^{r} t_i + (n - r)t$ is given in advance. More precisely, the test is continued until the total time on test T reaches a predetermined value T_0. If all sample elements fail before, the test terminates at the instant of the last (nth) failure ($r = n$).

In the simplest case when the probability of the event $r = n$ is negligibly small r is Poisson-distributed. However, even in the more general case an unbiased estimator and an exact confidence interval can be given. The unbiased estimator is:

$$\hat{\lambda} = \begin{cases} \dfrac{r}{T_0}, & \text{if} \quad r < n \\ \dfrac{n-1}{T}, & \text{if} \quad r = n \end{cases}$$

where r denotes the number of failures and T the total time on test until the termination.

Under the validity of the Poisson approximation, $D^2(\hat{\lambda}) \sim r/T_0^2$. For the more accurate formula, see [A 19], p. 180.

A confidence interval of level $1 - \varepsilon$ can be given in the following way:

If $r < n$ the confidence limits are

$$\frac{\chi^2_{\varepsilon/2,\, 2r}}{2T_0}, \quad \frac{\chi^2_{1-\varepsilon/2,\, 2(r+1)}}{2T_0}.$$

In the case $r = n$ the confidence limits are

$$\frac{\chi^2_{\varepsilon/2,\, 2n}}{2T}, \quad \frac{\chi^2_{1-\varepsilon/2,\, 2n}}{2T}.$$

In the cases of both (c) and (d), easy modification of the formula for confidence interval leads to critical values of a statistical test on λ (see case (a)).

If we wish to test the null hypothesis H_0: $\lambda = \lambda_0$ against the alternative $\lambda > \lambda_0$, the procedures corresponding to (c) and (d) may be suitably combined in the following way: We choose a pair of values r_0 and T_0 in such a way that

$$2\lambda_0 T_0 = \chi^2_{\alpha,\, 2r_0}$$

and the life test is continued until either r reaches r_0 or T (total time on test) reaches T_0. If the first of these two events take place first, the hypothesis will be rejected; otherwise it is accepted.

(e) Let us return to the case (b). If the investigation time t is fixed, and we wish to test the null hypothesis H: $\lambda = \lambda_0$ against the alternative $\lambda > \lambda_0$, the following (inefficient) test is usually used: we reject H_0 if $B(r \mid n, e^{-\lambda_0 t}) \geqslant 1 - \alpha$ and otherwise we accept it. Here $B(x \mid n, p)$ means the binomial probability distribution function

$$B(x \mid n, p) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i} \quad (x = 0, 1, \ldots, n).$$

If r/n is small, this test leads nearly to the same result as the test constructed on the Neyman–Pearson principle (see formulae (4.3.20–23) and the footnote after (4.3.23) in [A 19]).

The reliability may be estimated in the same way as in the case of (a). Alternatively, we may use an unbiased estimator [B 69].

Let us remark here that a distribution-free unbiased estimator for the reliability always exists, namely $\hat{R}(t) = r/n$ (see (j) below).

(f) *Exponential distribution; life test with replacement.* Now suppose that n units are placed on test and that every item that fails on test is immediately replaced. In this case the maximum-likelihood estimator for the hazard rate λ is

$$\hat{\lambda} = \frac{r}{nt}$$

where r denotes the number of failures occurring during the fixed test period t. It is unbiased.

$nt\hat{\lambda} = r$ is Poisson-distributed so that

$$D^2(\hat{\lambda}) = \frac{\lambda}{nt}\,.$$

A confidence interval of level $1 - \varepsilon$ is the following:

$$\frac{\chi^2_{\varepsilon/2,\, 2r}}{2nt} < \lambda < \frac{\chi^2_{1-\varepsilon/2,\, 2(r+1)}}{2nt}\,.$$

If, instead of t, r is predetermined, then nt is Gamma-distributed and the formulas will be slightly modified.

(g) *Exponential distribution; sequential test.* See Section 2.4.7.

(h) *Sampling tables for exponential distribution.* Sampling plans based on the above principles, including sequential sampling plans have been constructed and published [D 3, D 7]. The plans are classified according to three types: 1. life tests terminated upon occurrence of a preassigned number of failures, 2. life tests terminated at a preassigned time, 3. sequential life-testing plans (cf. II.2.4.7).

(i) *Weibull distribution.* The main advantage of the Weibull distribution lies in the easy graphical method by which the scale and shape parameters can be simultaneously estimated. This is performed by using the Weibull probability paper. The method is essentially the same as the estimation of the parameters of the normal distribution by means of the normal probability paper. (See Fig. 60 and the example on p. 371.)

Alternatively, the same estimation problem can be solved by more exact methods. The maximum-likelihood estimators for the parameters α and β of the distribution

$$F(x) = \begin{cases} 1 - e^{-x^{\beta}/\alpha}, & \text{if} \quad x \geqslant 0 \\ 0, & \text{if} \quad x < 0 \end{cases}$$

cen be obtained by solving the system of equations

$$n\alpha - \sum_{i=1}^{n} t_i^{\beta} = 0 ,$$

$$\frac{n}{\beta} + \sum_{i=1}^{n} \ln t_i - \frac{1}{\alpha} \sum_{i=1}^{n} t_i^{\beta} \ln t_i = 0 ,$$

where $t_1, t_2, \ldots, t_n$ are the observed failure times of the units. The numerical solution can be found by an iteration procedure.

The simultaneous estimation of two parameters is of doubtful practical value if the number of the observed units is small.

If the shape parameter is assumed to be known, statistical problems involving the scale parameter can be reduced to the case of exponential distribution. For this purpose, the transformation $t' = t^{\beta}$ is to be performed first.

Several other estimation procedures have been worked out [B 48]. These include the case of three-parameter Weibull distribution.

Sampling tables are published in [D 8–9].

(j) *Parameter-free procedures.* The reliability $R(t)$ can be estimated by the unbiased estimator r/n, where n is the sample size and r the number of failures prior to the time t.

TABLE 32

Time (100 hours)	Surviving units	Empirical distribution function (%)
0	200	0
1	197	1.5
2	190	5
3	181	9.5
5	146	27
7	101	49.5
10	23	88.5
20	1	99.5
30	0	100

For estimating the hazard rate function $\lambda(t)$, the estimator

$$\frac{n(w_1) - n(w_2)}{\sum_{i=1}^{r} (t_i - w_1) + n(w_2)\,(w_2 - w_1)}$$

may be applied. Here $n(x)$ is the number of units still working at the time point x_1 (so that $r = n(w_2) - n(w_1)$) and $t_1, t_2, \ldots, t_r$ are the failure times of the units that failed during the time interval (w_1, w_2); $w_1 \leqslant t \leqslant w_2$. The choice of the interval (w_1, w_2) is a matter of compromise. On the one hand, the interval must be sufficiently narrow for $\lambda(x)$ to be approximately constant; on the other hand, the standard error of the estimator calls for a long interval.

If $t_1, t_2, \ldots, t_r$ are not known, the estimators

$$\frac{\ln n(\omega_1) - \ln n(w_2)}{w_2 - w_1}$$

or

$$\frac{n(w_1) - n(w_2)}{(w_2 - w_1)\, n(w_1)}$$

may be applied.

A graphical estimator may be obtained by plotting $-\ln n(x)$ against x, drawing a smooth curve through the points, and determining its slope at the point t.

For these and other estimators, see [B 83].

Example: 200 units have been placed on life test. The first column of Table 32 gives the time points at which they have been controlled and the second column the corresponding numbers of surviving units. Assuming that the data come from a Weibull distribution, provide estimates of its parameters, by using Weibull probability paper.

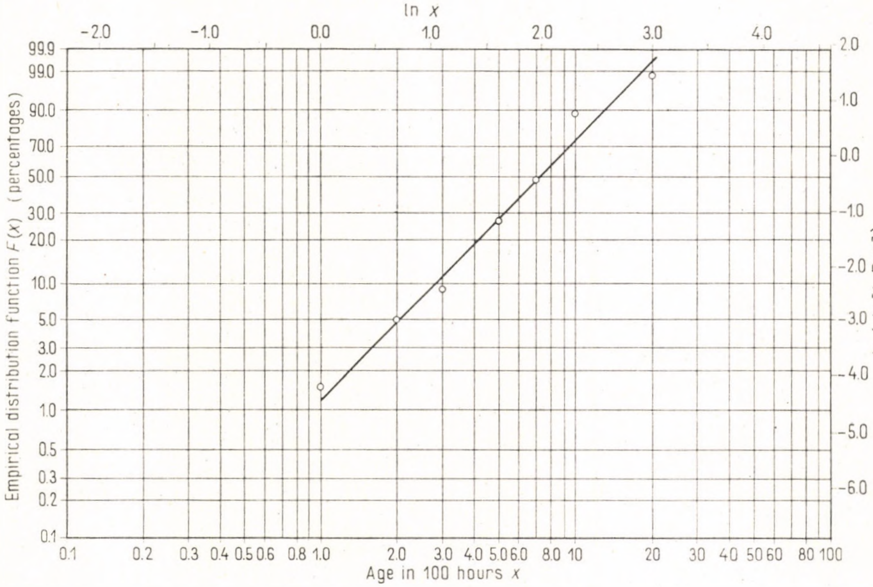


FIG. 60. Graphical estimation of Weibull parameters

The resulting points are approximated by a straight line (Fig. 60). The estimate for β is the slope of this line (difference of the ordinates corresponding to the abscissae 1 and 0): $\hat{\beta} = -2.4 + 4.4 = 2.0$ and $\ln \hat{a}$ (where $\hat{a}$ is the estimate for α) is the ordinate in the point 0: $\ln \hat{a} = -4.4$ hence $\hat{a} = .012$.

3.3 PROBLEMS RELATING TO COMPLEX EQUIPMENT

If a system or equipment is an assembly of components whose failure times are independent random variables, and if failure of any component causes the failure of the whole system (series arrangement), then the system hazard rate function is the sum of the component hazard rate functions:

$$\Lambda(t) = \lambda_1(t) + \lambda_2(t) + \ldots + \lambda_m(t),$$

where $\Lambda(t)$ is the hazard rate function of the system and $\lambda_1(t), \ldots \lambda_m(t)$ denote those of the components.

For the reliability function of this system

$$R(t) = R_1(t) \ldots R_m(t)$$

holds, where $R(t), R_1(t), \ldots, R_m(t)$ are the corresponding reliability functions.

Usually the exact values of the component characteristics are not known and, therefore, our formulae yield estimators subject to random errors. The estimation of their standard errors may be of interest and can be calculated.

For example, suppose that a system has m types of components, each with an exponential failure distribution, the numbers of the different components being $f_1, f_2, \ldots, f_m$, and the component hazard rate estimators obtained $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m$. Then the estimator of the system hazard rate is

$$\hat{\Lambda} = \sum_{i=1}^{m} f_i \hat{\lambda}_i$$

and its variance is

$$D^2(\hat{\Lambda}) = \sum_{i=1}^{m} f_i^2 D^2(\hat{\lambda}_i).$$

A considerable bias may occur if the hazard rate or other characteristic is estimated from component estimators which have been previously used for selecting the components from an assortment of competing products. This problem is dealt with in [B 66].

The exponential distribution plays a prominent role in models describing complex systems with renewal. Suppose the number of components is large, the failure times of the components are independently distributed and each failed item will be replaced by a similar one.

(a) It has been proved that under some mild conditions the failure distribution of the systems tends to an exponential distribution if both the number of components and the operation time tend to infinity. For a more detailed discussion, see, e.g., [A 2], pp. 18–21.

(b) Experience shows that in general the system failure distribution may be approximated by the exponential law but with a variable parameter depending on the age of the system.

The reasons for this are the following:

In practice, the hazard rate functions of the individual components are sufficiently smooth. The practical range of the system failure distribution is essentially smaller than the mean failure time of an individual component. The variation of the component hazard rate is negligible within such a short period. Therefore, this hazard rate may be considered constant from the point of view of the system failure distribution.

The parameters of the exponential distributions approximating consecutive times-between-failure are practically the same: although the hazard rate of the new component may differ from the old one, this has but very little effect on the value of the system hazard rate, the latter being the sum of the hazard rates of all components. For a longer period, this hazard rate varies slowly until the state mentioned in (a) is reached.

Most of the models constructed for complex systems assume that the failure times of the components are independent random variables. In practice, however, dependence between the failures of different components is frequently present. Multivariate models have been treated in the paper [B 26].

An important characteristic for equipment or systems in which failures are repaired is the *availability index*. This is the probability that under stationary conditions the system is working at a randomly selected time. The availability index is

$$A = \frac{\Theta}{\Theta + \Theta_r},$$

where Θ denotes the mean time between failures and Θ_r is the expectation of the repair time per failure (if repair is not started immediately after the failure, then Θ_r is to be interpreted as including the expectation of the waiting time as well).

There is a large variety of problems connected with calculations relating to systems constructed according to the redundancy principle. For example, let us consider the parallel redundancy system in which several subsystems are put into operation simultaneously, and any of them is capable of handling the job if failures occur in the other subsystems. In this case the overall reliability R of the whole system is

$$R = 1 - \prod_{i=1}^{m} R_i ,$$

where $R_1, R_2, \ldots, R_m$ are the reliabilities of the individual subsystems. (Independence of the subsystems is assumed.) The formula can be applied for time-dependent reliability functions, provided that the failure probabilities are the same if the subsystem is in action or is put out of work. Also, the same relation holds between the availability indices of the whole system and the subsystems.

Generally, the principles of parallel and series arrangements are combined in the construction of redundant systems.

The special problems of relay circuit systems will be dealt with in Section 3.4.

3.4 MULTICOMPONENT STRUCTURES

The models of the preceding Section assumed that systems and components were subject to one kind of failure only. In many instances, however, such models are not adequate for solving the relevant practical problems.

A typical example of failures which may take on either one of two forms is a two-terminal electrical network with contacts (mostly relays) as its components. Both the network and its components may meet two kinds of failure: failure to close and failure to open. Similar problems arise in connection with some non-circuit systems and their components, the functioning of which may be desirable or non-desirable at given instances.

The possibility of increasing the system reliability by means of redundant components modifies somewhat the problems which occur in systems such as those described in the preceding Section.

J. von Neumann [B 52] was the first to point to the fact that reliable systems can be constructed of relatively unreliable components if an appropriate redundancy is built in. Here, however, principles other than parallel redundancy must be applied in order to achieve decreasing probabilities of both kinds of failure. A more general theory of systems of this type has been developed by Moore and Shannon [B 51], and Birnbaum, Esary and Saunders [B 10].

We suppose now, as before, that the states of the system components uniquely determine the state of the whole system (network). The systems may be characterized by their *structure functions* and *structure reliability functions*.

Let the system have n components and let x_i be the indicator variable of the ith component, i.e., $x_i = 1$ if the ith component functions (establishes contact) and $x_i = 0$ otherwise. Let $\varphi(x_1, x_2, \ldots, x_n) = \varphi(x)$ be a function which takes on the value 1 if the system functions, and 0 otherwise. Here x denotes the vector $[x_1, x_2, \ldots x_n]$ and $\varphi(x)$ is called the *structure function.*

Now let us suppose that the quantities x_i are independent random variables with $P(x_i = 1) = p_i$, then the function

$$h(p_1, p_2, \ldots, p_n) = E(\varphi(x))$$

is the *structure reliability function*. Often it is assumed that $p_1 = p_2 = \ldots = p_n = p$ and we write simply $h(p)$ instead of $h(p, p, \ldots p)$.

Now let $p_{(2)}$ and $p_{(1)}$ denote the values of p when the system is desired to operate (switched on), and not to operate, respectively. Clearly, it is desirable to construct a system with $h(p_{(2)})$ as near to 1 as possible, and at the same time $h(p_{(1)})$ as small as possible.

Let us remark that not all possible structure functions are physically realizable by a circuit network. If $\varphi(x)$ is physically realizable by a two-terminal circuit network, then the corresponding network or system is called *simple network structure*.

Moore and Shannon [B 51] have proved that if $h(p)$ is the structure reliability function of a simple network structure, $h(p) \not\equiv p$, and $h(p_0) = p_0$ for some value of p_0, $0 < p_0 < 1$, then $h(p) < p$ for $0 < p < p_0$ and $h(p) > p$ for $p_0 < p < 1$. In other words, in the interval $(0, 1)$, $h(p)$ intersects the straight line through the origin of slope 1 at most once, and if it intersects, then from below. Moreover, they have given a lower bound for the number of components necessary for assuring $h(p_{(1)}) < \varepsilon_1$ and $h(p_{(2)}) > 1 - \varepsilon_2$ (for given $p_{(1)}, p_{(2)}, \varepsilon_1, \varepsilon_2$).

Birnbaum, Esary and Saunders have extended these results for a larger class of systems: the *coherent structures* (*monotonic structures*) for which the corresponding structure functions $\varphi(x_1, x_2, \ldots, x_n)$ are monotone increasing (nondecreasing) in all of their arguments, the trivial cases $\varphi(x_1, x_2, \ldots, x_n) \equiv 0$ and $\varphi(x_1, x_2, \ldots, x_n) \equiv 1$ being excluded. In other words, coherent structures have the property that replacing a defective element by a good one cannot result in the failure of a functioning system. Clearly, simple network structures possess this property, i.e., the class of simple network structures is a subclass of the class of coherent structures.

Birnbaum, Esary, and Saunders [B 10] have proved that if a coherent structure reliability function intersects the line with slope 1 it does so only once and from below (i.e., $h(p)$ is S-shaped). This result has been extended in a certain sense by Esary and Proschan [B 25], for the case when the probabilities $p_i = P(x_i = 1)$ are different.

An important subclass of the class of coherent structures is that of the *k-out-of-n structures.* A k-out-of-n structure is a system with n components, the system functioning if and only if at least k of its components are functioning $(1 \leqslant k \leqslant n)$. The 1-out-of-$n$ structures are the parallel structures, the n-out-of-n structures are the series structures.

It was shown by Birnbaum, Esary and Saunders that the k-out-of-n structures are, in a certain sense, optimal within the class of coherent structures. More precisely, if $h(p;\ k, n)$ denotes the structure reliability function of the k-out-of-n structure and $h(p)$ the structure reliability function of any other system containing n components, then if $h(p;\ k, n)$ intersects $h(p)$ in $(0, 1)$, it does so only once and from below.

3.5 CLASSES OF MODELS FOR FAILURE DISTRIBUTIONS BASED ON THE PROPERTIES OF THE HAZARD RATE FUNCTION

As noted in Section 3.1, in most practical cases the hazard rate is a piecewise monotone function: in the "debugging" period it is decreasing, then it remains constant for some time, and finally it increases due to "wear-out". Therefore, classes of distributions with monotone hazard rate—and, more generally, classes based on some special properties of the hazard rate function—play an important role in reliability theory.

The theory of distributions with monotone hazard rate has been developed mostly by Barlow and Marshall [B 2]. We use here their abbreviations: IFR for increasing and DFR for decreasing hazard rate, (hazard rate = failure rate) and for the corresponding distributions as well. Note that the terms "increasing" and "decreasing" are used in the weak sense, consequently, the exponential distribution belongs to both classes.

The definition of the hazard rate and consequently that of the classes just mentioned may be extended to the case of discrete distributions as well; nevertheless, we will deal here with continuous distributions only.

The above-mentioned classes have been proved to possess the following preservation properties:

(1) If $F_1(x)$ and $F_2(x)$ are IFR, then their convolution

$$H(t) = \int_{-\infty}^{\infty} F_1(x - y) f_2(y) dy$$

is also IFR. Here $f_2(x) = F_2'(x)$. Consequently, a system consisting of a single unit supported by $n - 1$ reserve units is IFR if all components are IFR.

(2) The order statistics of an IFR failure distribution are IFR as well. This property assures that the k-out-of-n structures (see Section 3.4) consisting of n identical IFR components will also be IFR.

(3) A mixture of DFR distributions is DFR, i.e., if $F_1(x)$ is DFR, $a_i \geqslant 0$ $(i = 1, 2, \ldots)$ and $\sum_{i=1}^{\infty} a_i = 1$, then

$$G(x) = \sum_{i=1}^{\infty} a_i F_i(x)$$

is a DFR distribution.

For the proofs and for further similar properties of IFR distributions see [A 2], pp. 35–39.

Many inequalities have been proved for distributions of this type. Let us mention the following:

(1) If $F(x)$ is IFR with expectation μ, then

$$1 - F(t) \geqslant e^{-t/\mu}, \text{ if } t < \mu$$

$$1 - F(t) \leqslant e^{-\omega t}, \text{ if } t > \mu$$

where ω is defined by $1 - \omega\mu = e^{-\omega t}$.

(2) If $F(x)$ is DFR with expectation μ then

$$1 - F(t) \leqslant \begin{cases} e^{-t/\mu}, & \text{if } t \leqslant \mu \\ \mu e^{-1}/t, & \text{if } t \geqslant \mu. \end{cases}$$

For the proofs, further inequalities, and tables see [A 2], pp. 26–45, 236–240 and [B 2].

Another important class of distributions is the class with increasing hazard rate on the average (IFRA), i.e., distributions, for which $\int_0^x \lambda(t)\, dt/x$ is nondecreasing for $x > 0$. ($\lambda(x)$ is the hazard rate function.) As Birnbaum, Esary and Marshall [B 10] have shown, this is the smallest family of failure distributions containing the exponential distribution which is closed under the formation of coherent structures and limits in distribution.

Some other classes of distributions have been investigated by Saunders [B 70].

A number of tests for exponentiality versus IFR and IFRA has been investigated by various authors (see [B 4]). For example, the test based on the "total time on test" statistic

$$\frac{\sum_{i=1}^{n-1} (n-i)\, t_i^*}{\sum_{i=1}^{n} t_i}$$

where $t_1, t_2, \ldots, t_n$ are the sample elements, and $t_1^*, t_2^*, \ldots, t_n^*$ the order statistics, turns out to have advantageous properties.

APPENDIX

SPECIAL TABLES

TABLE I

RANDOM NUMBERS FROM THE UNIFORM DISTRIBUTION

No. of row	Number of column										No. of row
	1	2	3	4	5	6	7	8	9	10	
1	69 89	48 66	79 43	19 70	22 05	07 65	88 02	21 76	66 04	62 11	1
2	24 77	94 57	87 60	07 16	40 45	06 45	57 65	40 15	02 82	05 35	2
3	72 65	88 34	62 93	90 93	70 66	70 05	67 47	22 08	61 65	09 68	3
4	39 30	10 50	42 82	69 89	61 34	34 68	22 80	72 17	39 39	59 98	4
5	09 11	14 62	29 40	57 39	22 73	06 70	25 06	04 33	69 63	16 12	5
6	51 11	11 12	88 81	76 94	19 37	13 35	63 47	42 08	41 84	81 38	6
7	65 77	58 61	89 64	63 39	88 71	42 60	28 82	11 20	28 53	80 97	7
8	79 12	14 66	97 10	84 56	96 08	10 09	12 86	89 73	55 76	95 59	8
9	38 75	48 64	50 17	38 86	77 39	74 48	03 03	34 38	66 39	81 83	9
10	58 40	22 21	22 67	46 62	25 00	49 26	15 60	97 55	23 01	03 61	10
11	22 22	46 73	60 94	27 99	24 32	06 19	61 03	83 69	72 81	76 57	11
12	02 71	34 38	81 36	61 05	18 17	48 67	64 73	09 75	02 70	04 52	12
13	92 39	36 71	26 70	79 42	53 64	72 90	91 71	52 25	92 48	19 47	13
14	40 79	82 43	03 49	12 21	15 41	93 86	55 99	24 32	22 43	99 83	14
15	38 20	54 22	04 00	39 43	60 30	28 56	97 02	17 83	96 64	72 08	15
16	01 47	69 71	21 02	12 08	54 41	40 36	74 91	15 78	24 38	32 13	16
17	17 71	98 82	87 19	98 06	23 86	46 69	47 99	22 62	78 63	37 25	17
18	14 40	28 22	68 48	42 01	15 09	20 10	82 06	68 16	03 89	52 54	18
19	61 71	92 22	42 33	98 46	21 13	98 20	56 60	45 43	48 76	93 08	19
20	25 91	04 91	31 66	54 66	46 88	26 99	84 63	51 72	37 09	57 06	20
21	62 33	60 61	25 68	89 31	71 05	69 55	17 81	63 69	60 52	58 24	21
22	25 99	51 07	10 19	91 59	64 40	22 13	06 69	82 39	48 30	35 73	22
23	67 73	62 18	96 07	22 70	47 08	09 82	38 31	05 40	21 88	30 25	23
24	18 42	10 91	64 09	09 09	59 65	98 55	38 99	46 15	61 28	89 76	24
25	12 59	48 14	79 58	98 77	35 21	36 84	79 02	11 89	22 10	96 69	25
26	30 97	09 57	23 93	57 59	52 43	80 91	33 13	34 65	54 47	66 96	26
27	24 03	63 23	59 65	01 82	75 54	08 64	49 21	90 44	68 39	26 26	27
28	17 71	81 75	79 97	71 24	34 76	29 10	54 12	64 57	60 12	91 36	28
29	84 78	14 99	72 29	79 74	98 70	85 92	86 06	03 15	13 73	69 78	29
30	90 42	72 99	11 51	17 71	40 16	35 07	55 98	71 03	59 52	11 33	30
31	82 83	28 20	58 23	72 28	89 20	43 05	89 54	07 55	36 59	60 97	31
32	22 60	95 45	37 34	57 92	68 33	93 94	37 24	12 81	31 97	61 82	32
33	86 36	28 34	74 77	71 72	09 78	38 96	92 86	26 52	62 06	39 08	33
34	18 56	85 33	92 06	64 25	76 61	15 55	83 39	49 10	77 11	11 44	34
35	03 76	56 95	85 35	33 30	57 94	02 95	70 16	03 81	42 58	26 24	35
36	46 17	98 63	44 54	13 27	51 05	55 78	08 21	17 72	67 86	92 36	36
37	51 60	75 31	81 30	10 19	90 28	50 58	80 28	93 18	79 84	97 77	37
38	80 55	53 75	59 64	73 62	74 95	45 69	27 16	45 03	94 93	56 28	38
39	27 16	59 02	27 72	23 20	57 68	33 14	58 26	64 86	12 94	86 79	39
40	53 02	61 94	09 60	59 22	89 42	04 57	53 39	35 30	88 35	79 15	40
41	76 82	21 39	36 50	56 76	58 75	28 80	52 66	60 76	94 61	19 52	41
42	59 63	34 80	51 01	74 75	09 44	97 55	85 19	88 33	61 46	53 85	42
43	84 23	39 26	03 39	60 86	59 94	17 12	08 10	13 90	71 35	29 13	43
44	61 18	53 86	89 45	51 09	60 11	97 58	37 75	76 43	01 63	40 26	44
45	34 00	85 00	81 36	41 48	02 87	41 77	89 48	70 09	21 10	07 42	45
46	05 92	95 30	98 81	48 27	85 92	76 65	44 55	36 13	60 12	75 12	46
47	66 39	34 32	44 07	13 24	63 15	17 94	29 80	03 86	45 87	29 51	47
48	69 21	92 44	29 73	45 12	30 40	61 37	96 33	47 75	22 55	76 88	48
49	94 59	10 31	91 09	63 16	38 80	02 93	15 83	21 66	15 63	74 77	49
50	15 43	99 00	52 47	43 63	63 57	38 27	52 29	49 87	80 74	60 27	50

Example: From a lot of size 800, its pieces labelled from 1 to 800, a sample of size 75 is to be taken. We begin with the arbitrary choice of a cell (e.g., 14th row, 2nd column), where the reading of the number will start. Successively the first (or last) three figures from each four-digit number will be regarded. Those which

TABLE I (cont.)

No. of row	Number of column										No. of row
	1	2	3	4	5	6	7	8	9	10	
51	25 99	09 13	49 90	47 27	07 24	01 16	01 44	21 30	92 54	42 45	51
52	90 35	68 61	78 74	63 45	13 23	76 84	61 81	66 46	34 62	67 48	52
53	03 91	17 33	91 45	39 41	44 76	50 66	06 35	42 36	73 82	11 53	53
54	35 36	29 20	64 43	55 51	07 60	67 36	29 81	67 08	76 19	51 16	54
55	42 22	49 22	92 50	12 58	86 52	40 48	13 82	55 83	47 63	70 04	55
56	63 13	62 74	99 82	30 47	91 70	17 46	61 20	91 08	71 91	02 31	56
57	38 26	86 35	68 65	88 37	21 71	23 26	70 68	97 46	08 98	74 85	57
58	07 03	88 07	39 40	44 26	68 58	80 03	03 60	14 94	64 88	08 38	58
59	26 97	65 49	53 54	18 01	77 95	81 36	74 91	23 07	88 98	78 22	59
60	36 60	22 52	81 20	13 09	79 96	38 46	44 01	76 82	15 77	36 81	60
61	93 53	71 50	76 35	73 19	27 29	00 72	49 48	91 91	14 21	33 32	61
62	46 33	52 51	29 86	83 10	49 44	12 33	25 73	61 28	45 03	96 84	62
63	13 63	92 29	70 92	12 69	58 84	24 88	57 99	27 08	40 05	45 41	63
64	28 76	97 05	39 74	07 91	38 43	43 40	12 56	29 24	61 92	04 86	64
65	13 73	64 83	47 68	86 89	92 41	98 20	92 00	36 39	20 11	00 58	65
66	26 94	44 01	57 01	42 45	22 10	45 12	22 96	63 68	22 04	63 36	66
67	69 51	45 93	48 47	17 30	33 20	46 46	37 81	99 50	61 28	94 85	67
68	63 27	57 33	23 65	00 24	96 77	82 87	06 33	10 22	24 98	92 57	68
69	89 93	51 85	83 29	83 54	23 83	54 15	17 80	82 64	87 34	08 02	69
70	86 00	65 70	21 80	76 77	73 28	69 24	76 73	33 27	85 97	01 97	70
71	72 73	52 38	77 75	84 71	24 22	36 16	59 86	18 05	75 10	09 03	71
72	88 81	04 27	87 93	19 44	78 55	79 84	66 80	75 85	18 18	21 20	72
73	24 87	79 26	27 86	62 12	03 96	59 55	09 14	78 21	44 98	47 49	73
74	00 78	29 22	99 98	19 95	30 10	57 11	46 94	53 69	80 62	70 13	74
75	13 41	60 86	67 66	79 24	39 75	39 13	09 02	93 37	85 08	78 83	75
76	52 37	70 26	99 70	30 36	38 07	74 87	32 51	30 95	05 39	61 96	76
77	75 78	13 65	82 04	20 11	98 55	32 03	74 46	62 90	38 53	36 23	77
78	31 06	55 86	99 17	18 04	17 13	19 52	98 82	95 74	10 16	70 06	78
79	95 70	48 50	25 75	21 82	85 42	67 34	37 53	67 94	51 72	10 21	79
80	03 27	99 08	18 21	06 26	85 54	04 73	39 61	32 49	34 75	53 18	80
81	64 06	36 20	60 01	32 42	02 56	28 50	19 88	25 91	86 24	14 81	81
82	75 43	68 01	77 69	33 11	53 71	23 27	15 44	13 74	74 99	68 48	82
83	50 34	70 26	04 01	04 09	66 98	72 36	11 83	47 70	08 88	92 85	83
84	85 07	09 03	78 46	78 23	78 97	83 14	14 15	61 78	71 37	56 79	84
85	42 28	33 82	74 87	34 77	98 07	52 11	70 24	29 86	19 30	30 58	85
86	14 92	47 64	84 53	31 23	54 15	21 85	44 00	14 44	68 71	58 31	86
87	28 88	41 02	14 50	17 05	44 55	10 40	75 52	83 99	74 99	22 03	87
88	24 55	43 31	73 46	99 35	57 73	48 31	78 39	93 62	10 91	77 44	88
89	85 68	03 72	36 59	50 85	10 58	03 94	16 57	91 44	32 25	25 12	89
90	91 11	35 70	79 28	08 36	87 17	82 20	13 80	74 17	93 66	85 45	90
91	97 71	60 96	83 77	71 69	24 57	44 04	12 30	57 30	14 38	58 38	91
92	46 03	49 06	19 41	79 79	50 05	59 50	61 60	92 45	90 06	91 63	92
93	30 51	69 57	24 33	28 35	35 90	70 35	93 79	55 24	57 47	92 75	93
94	11 19	36 27	13 37	27 45	75 88	31 77	07 31	18 82	23 62	55 43	94
95	16 92	24 12	83 68	18 40	52 69	53 91	28 31	78 90	89 01	49 24	95
96	69 08	34 07	09 53	37 71	81 75	61 86	40 47	36 55	30 90	45 00	96
97	75 52	73 31	35 42	93 21	80 95	35 88	46 48	27 66	20 70	19 88	97
98	30 73	65 94	76 54	30 82	81 46	20 08	33 90	67 96	21 03	08 76	98
99	77 12	73 30	48 24	40 74	86 90	80 68	93 75	45 79	17 40	24 15	99
100	55 00	78 16	05 41	51 15	96 40	43 34	08 45	03 77	43 25	26 20	100

exceed 800 or which have already occurred, will be omitted and the others yield the random sample; in our example thus we obtain: 243, 349, 221, 541, 386, 599, 432, (243), (983), (820), 422, 400, etc.

Table II

Random numbers from the standard normal distribution

	1	2	3	4	5	6	7	8	9	10
1	0.9488	−0.4825	0.4652	−0.8352	−1.7609	−0.8430	0.3427	−0.4233	1.0126	−0.7874
2	1.1280	1.5796	−0.7403	−0.8748	0.2283	0.5984	−0.8884	2.0270	−0.8384	−1.1196
3	−0.2625	1.0034	−0.7292	−1.8604	−2.9435	−0.0032	1.1363	0.5187	0.6983	−0.5943
4	−0.9451	−1.1655	−2.7899	1.7438	0.3809	−1.6274	0.8006	0.2146	1.2386	0.2793
5	−1.3999	−0.7399	−0.8896	−0.5806	0.1512	−0.9559	0.3845	0.8279	−0.6445	0.1616
6	1.7036	−1.6287	−0.6434	0.1521	−0.3988	−0.5167	−1.2492	0.1219	−0.9480	0.2038
7	0.8439	−0.2719	−1.3963	1.0963	−1.0114	0.7661	−0.2078	−0.8741	−0.6470	−0.6921
8	0.2522	−1.0542	−0.6106	−0.5306	0.4809	−0.8280	1.0427	0.2759	−0.0101	−0.3880
9	−1.5623	−1.2316	−0.0172	0.2178	0.1304	1.3850	0.9719	1.0907	0.0327	2.1343
10	−0.0006	0.9479	−0.2996	−1.1899	1.7240	−0.0482	0.1893	−0.8962	−0.5009	−0.6823
11	−1.4171	0.4747	−1.6690	0.7354	−0.9964	−0.6844	0.0449	1.4389	−0.6187	1.2785
12	−0.3107	1.6957	−0.4997	2.0610	−0.4609	−0.3938	−1.4355	−1.7797	0.5167	−0.8751
13	0.0206	−1.6674	2.6840	1.6944	1.3848	0.8773	0.3288	1.7371	0.0952	2.1660
14	0.7118	−0.9071	0.1369	−0.2340	−2.1880	1.1900	0.2918	1.1753	−0.1624	1.6286
15	−0.4218	0.3833	−0.9962	2.3147	0.0482	−0.4741	−0.2161	−1.3685	0.5275	1.3641
16	−1.1624	−0.3455	0.2122	−1.8386	−0.3436	−0.1025	−1.2876	0.6254	−0.2053	0.9644
17	−0.8462	−0.1206	−1.5406	1.1488	1.2651	−2.0438	−1.2225	0.7262	−0.2877	0.2200
18	0.1232	1.2594	−1.2022	−1.1104	−2.8275	0.4543	1.1887	0.3957	−1.5398	−1.2631
19	−2.0216	−0.2612	−0.3534	−0.5621	1.1708	0.8508	0.3556	−0.4923	1.4284	−1.8755
20	0.1623	−0.0081	−2.1268	1.5494	0.7871	−1.0936	0.7635	0.4532	0.3208	0.5561
21	−0.7950	−0.2277	−0.2613	−0.3327	−0.3832	0.5205	1.2046	0.7529	−1.4660	0.6500
22	−0.3054	−0.8586	0.4165	1.9017	−2.5146	−1.0833	0.9828	−1.9563	−0.5216	0.3566
23	0.3027	0.6457	−0.7764	1.3396	−0.7356	−1.0909	−0.2543	−0.2580	−1.1195	−0.7685
24	0.0577	−1.7205	−0.6596	1.3489	−1.2487	0.5882	−1.5674	0.8480	2.3988	−0.8677
25	0.7504	0.4703	−0.2526	0.4382	0.7306	0.7379	0.5939	−2.1859	1.0300	0.7261
26	−1.9212	0.3372	−1.3168	−0.4175	1.0473	−0.0360	0.2945	0.6970	1.1300	−1.0122
27	−0.0924	−1.7182	0.4619	−0.3742	1.6348	0.3444	1.6214	−0.1907	0.6932	−0.0810
28	−0.8591	−0.5374	−0.3420	1.3310	−0.6769	−0.1345	1.4955	0.0418	−0.0804	−0.0841
29	−0.0773	0.3093	−1.4594	0.2876	−0.4421	1.1403	−1.1846	0.1985	−1.5202	0.4567
30	−1.4914	1.1044	1.1166	0.4313	−0.7245	−0.8845	0.1591	−0.7299	2.3509	0.8032
31	−0.5588	0.6305	−0.5034	−0.2746	−0.0495	−0.8129	1.0247	−0.4111	0.9321	−0.7801
32	−0.0652	0.0536	−1.0630	−1.0671	0.2739	−0.9821	−0.8807	−0.6402	−0.6860	−0.9002
33	−0.6218	1.4251	−0.0892	−2.2451	0.2858	1.1588	0.5252	0.2272	−0.8517	−1.9208
34	1.9547	−0.1307	0.4914	−0.9649	0.3336	1.2052	−0.9182	−0.5294	−0.7226	0.4823
35	1.9498	0.4926	0.3567	−0.7205	2.7574	−0.1516	−0.0576	−0.0241	0.8167	−0.8946
36	1.0678	0.5369	0.1469	0.0499	−0.5118	−0.6062	−0.0309	−1.1272	0.4575	−1.0642
37	0.0883	−0.9095	2.0149	−0.6461	−1.0842	−0.1087	−1.0724	−0.5347	1.5288	−0.1792
38	0.1358	−1.0087	1.1364	0.4760	0.4278	−0.9738	−1.6150	−0.6007	−0.1491	2.2408
39	−2.2061	0.3789	−0.5341	1.4832	0.8861	0.2615	2.1069	−0.7424	−0.5356	0.3172
40	1.6095	0.2060	0.7259	0.0407	−1.7604	−0.6768	−0.8278	−0.4253	1.1536	0.7529
41	0.2065	0.2729	−1.8502	1.4658	−0.3645	1.1769	0.2204	0.9865	0.0907	0.8630
42	−0.5313	0.1152	−0.8527	−1.1293	0.1958	0.4263	−2.0932	−0.0361	0.4714	−0.9382
43	−0.1641	0.3072	0.0255	0.9264	−0.2014	1.4558	−0.0618	0.1628	−0.0953	0.9157
44	−0.3064	−0.1655	−2.7458	−1.2988	0.4283	0.1233	1.1311	−0.3308	−0.6206	−1.5758
45	−1.8309	0.2759	0.2515	0.0261	−1.0472	0.3694	0.3213	0.9057	0.1540	1.3660
46	−0.7503	2.8039	−0.3788	−0.5724	−0.0077	0.6109	0.3141	1.7401	1.5365	−0.6533
47	1.7913	−0.3481	−0.8560	0.0772	0.9704	1.4603	−0.0712	−0.1207	−0.0526	−1.0633
48	1.5554	−1.2221	2.3516	0.8071	−0.1665	0.3510	−1.8130	−2.0461	−1.2053	0.4230
49	−1.0651	−0.2187	−0.0169	0.7563	−0.9081	−0.6040	1.0019	−0.0741	0.0457	0.6000
50	−1.4538	−1.2840	−0.7511	0.8853	−1.7494	0.6990	0.4424	−1.5259	0.1111	−0.9760

Example: A type of equipment contains 20 pieces of a certain part. The life length Y of this part is distributed according to $N(\mu, \sigma)$. The question is the distribution of the breakdown of the instrument caused by the failure of any of these parts. The distribution of the smallest sample element from a sample of size 20 taken from $N(\mu, \sigma)$ is to be determined.—For this reason the numbers of the table will be divided into groups of twenty (e.g. according to the columns) and

TABLE II (cont.)

	1	2	3	4	5	6	7	8	9	10
51	0.1120	0.0744	—1.0336	—1.0587	—1.1827	—0.8379	—0.4996	0.3556	1.2233	0.1136
52	—0.4891	—0.3548	—1.5676	0.0019	—0.5767	—0.8709	—1.3858	0.4372	0.1731	0.7872
53	—0.2663	0.3007	1.5049	—0.9553	0.1764	—2.0669	—0.7475	—0.5945	—0.2246	1.9098
54	0.9902	—0.2221	—0.3399	—0.8119	—0.3304	—1.1286	—0.0519	—0.7439	—1.1035	0.6949
55	—2.0535	1.0965	—0.8173	1.7224	0.1700	—1.7641	0.3495	0.7687	—1.0858	—0.7107
56	0.6796	—0.4230	1.2926	1.8932	2.4078	0.0344	0.1516	—0.1813	1.2236	0.1069
57	0.0833	—1.0516	0.4341	—1.5289	0.8839	—0.9582	0.7686	0.8224	—0.9561	—0.1274
58	2.6771	1.2233	1.7631	0.3620	—1.0163	—0.9239	1.4274	—0.4993	—0.2873	—1.1683
59	0.9755	1.1832	—0.7584	—0.7593	—0.4783	—0.7923	—1.1528	1.0681	0.2419	0.6231
60	0.5690	—1.1152	2.0734	—1.9592	0.8448	0.3519	—1.2526	0.7210	—0.1936	1.1790
61	—0.9577	—1.7919	—0.6119	0.0120	—0.4284	0.0442	1.2660	—0.7814	0.7665	1.4127
62	2.9940	—0.0075	—0.1585	—0.0750	—0.3133	—0.4389	1.2060	—0.8645	0.4896	1.3489
63	0.9580	0.7375	1.3365	—0.9219	1.2584	2.0900	—2.3689	1.3264	—0.5792	—0.6092
64	—0.4356	—0.0025	0.0903	0.4218	—0.3885	0.4629	0.9825	1.2744	0.1332	—0.2091
65	0.6339	0.0469	—0.2643	—1.3297	1.7475	—0.4095	—0.3888	—0.5887	—0.8772	0.5724
66	—0.4956	—0.0221	0.8602	—0.4173	—2.3558	1.4228	—0.7755	—1.9336	1.1205	1.8741
67	0.4128	0.4908	—0.7834	—1.9835	—0.2022	—1.6233	0.8173	0.3037	—1.1510	—0.9682
68	—0.5204	—0.6121	1.4206	—0.2111	0.3496	1.1595	0.0290	—0.6472	0.0031	—0.5181
69	—0.0932	0.4330	—0.2089	—0.6982	—0.3518	—1.2791	—0.7194	—0.9866	0.5389	—1.6670
70	0.8607	—0.7345	—0.7715	0.0423	0.8904	1.4608	0.2882	—0.3760	—0.2051	—0.5454
71	—1.7076	—0.3933	0.6515	1.5048	—0.0730	—0.6638	0.4034	—1.8660	—1.2476	—0.0428
72	0.4928	0.8381	—0.3753	0.1144	—0.2182	—1.0529	—1.0286	0.5004	1.0291	0.3235
73	0.6895	0.4955	0.3328	0.5754	0.1975	—1.4052	—0.1710	—0.4187	—0.5333	—0.3920
74	1.4744	0.4648	—0.2868	1.3085	0.0781	—1.6609	—1.1381	2.3549	1.9089	1.1986
75	2.4836	2.0358	1.4443	0.1454	0.4472	0.7847	—1.5508	—0.1515	1.2931	—1.5741
76	0.0316	—0.3438	—0.2765	1.0870	—1.0676	1.4006	—0.2445	0.7056	—0.1010	0.9101
77	—0.0995	—0.3895	0.4220	—0.2234	—0.9696	—0.1764	—0.2497	—0.1858	1.0134	—0.8583
78	—0.0113	—0.6999	—0.0467	—1.0913	—2.5988	—0.5597	—1.7375	1.2671	0.9697	—0.1952
79	1.1628	0.7301	0.5449	—0.5353	2.0278	—1.3400	—0.1121	0.1789	1.7905	—1.4428
80	0.8340	—2.3824	—1.0465	1.1809	—0.8343	0.9328	—0.8685	2.3094	—0.3563	—0.1527
81	1.6132	1.2653	1.7987	—1.4090	—2.8804	—0.5221	0.7183	1.1580	1.0909	0.6534
82	0.2402	1.1960	—0.3200	1.4510	0.8075	—2.4153	0.8446	—0.0322	—0.5741	0.5284
83	0.7746	—0.5576	—1.2124	0.4770	—0.8087	—0.0967	—0.0326	0.9516	0.1681	—0.1038
84	0.9466	0.5345	—1.3366	0.1320	—0.1058	—0.1337	1.1456	0.7856	0.1346	0.7282
85	0.0832	0.7703	—0.4910	0.0890	1.1083	1.1489	1.2994	1.1045	1.9648	—2.3493
86	—1.0306	0.1139	—0.6877	0.2891	0.8411	0.9275	0.3203	—1.4504	0.8646	0.5181
87	0.6518	0.4150	0.1051	—0.0712	—0.1023	—0.4082	0.2543	—0.7463	0.5572	0.2129
88	—0.8598	—1.5648	0.0416	1.0001	—1.2453	—0.0100	—0.0465	—1.0903	0.9775	—1.3733
89	—0.1279	2.1924	0.5982	1.4295	—0.3406	1.3284	1.7133	1.0295	1.0734	0.1529
90	0.5486	—2.7312	1.2093	0.6238	0.8535	0.9189	—0.8862	0.4593	0.5149	—0.4742
91	—1.1572	—0.2530	0.9927	0.5860	0.5351	—0.3496	—0.3770	—1.1901	—1.0512	—0.0191
92	—0.3654	—0.2991	0.1143	1.1057	0.4526	—0.9011	—0.1870	—0.9302	—0.6318	—1.4360
93	1.2690	1.6686	—3.1223	1.1544	0.4415	0.2482	1.7710	0.3901	—0.9265	—0.3890
94	—0.4880	—0.8883	—0.2750	1.2774	—0.0967	2.4488	0.1233	0.6925	0.0557	—0.2133
95	1.7239	0.6339	—0.5750	0.7454	2.0000	0.7462	—0.9246	—1.1874	0.0736	0.2896
96	—0.4254	0.4050	0.1204	—0.6698	0.0302	0.6639	—0.7866	—0.4581	—1.4095	—1.7356
97	0.5588	—0.3525	0.1409	—0.9466	0.4077	0.3023	—0.9316	0.4861	—0.0984	—0.9593
98	—1.0448	—0.2256	0.1224	1.2097	—1.0597	—0.0691	—0.5130	0.6929	—0.1526	—0.2143
99	0.3032	0.7075	—2.2340	—1.1439	0.7687	0.3107	—1.2128	—1.0571	1.0247	—1.1101
100	—0.7191	—0.5151	—0.3727	—0.2902	—0.6939	—0.7956	1.0395	—0.7808	—0.9632	—0.6110

the smallest of each group will represent the smallest element of a sample of size 20. Thus from the 1000 random numbers 50 smallest element X_1^* will be obtained, and from the values $Y_1^* = \mu + \sigma X_1^*$ an empirical distribution function can be constructed. — If the distribution function of Y is $F(x)$ instead of $\Phi(x;\ \mu,\ \sigma)$, then the Y_1^* values can be obtained from the relation $Y_1^* = F^{-1}(\Phi(X_1^*))$.

TABLE III

VALUES OF THE DENSITY FUNCTION

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

OF THE STANDARD NORMAL VARIABLE

$(\varphi(-x) = \varphi(x))$

x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$
0.00	0.3989	0.52	0.3485	1.04	0.2323	1.56	0.1182	2.08	0.0459	2.60	0.0136
0.01	0.3989	0.53	0.3467	1.05	0.2299	1.57	0.1163	2.09	0.0449	2.61	0.0132
0.02	0.3989	0.54	0.3448	1.06	0.2275	1.58	0.1145	2.10	0.0440	2.62	0.0129
0.03	0.3988	0.55	0.3429	1.07	0.2251	1.59	0.1127	2.11	0.0431	2.63	0.0126
0.04	0.3986	0.56	0.3410	1.08	0.2227	1.60	0.1109	2.12	0.0422	2.64	0.0122
0.05	0.3984	0.57	0.3391	1.09	0.2203	1.61	0.1092	2.13	0.0413	2.65	0.0119
0.06	0.3982	0.58	0.3372	1.10	0.2179	1.62	0.1074	2.14	0.0404	2.66	0.0116
0.07	0.3980	0.59	0.3352	1.11	0.2155	1.63	0.1057	2.15	0.0396	2.67	0.0113
0.08	0.3977	0.60	0.3332	1.12	0.2131	1.64	0.1040	2.16	0.0387	2.68	0.0110
0.09	0.3973	0.61	0.3312	1.13	0.2107	1.65	0.1023	2.17	0.0379	2.69	0.0107
0.10	0.3970	0.62	0.3292	1.14	0.2083	1.66	0.1006	2.18	0.0371	2.70	0.0104
0.11	0.3965	0.63	0.3271	1.15	0.2059	1.67	0.0989	2.19	0.0363	2.71	0.0101
0.12	0.3961	0.64	0.3251	1.16	0.2036	1.68	0.0973	2.20	0.0355	2.72	0.0099
0.13	0.3956	0.65	0.3230	1.17	0.2012	1.69	0.0957	2.21	0.0347	2.73	0.0096
0.14	0.3951	0.66	0.3209	1.18	0.1989	1.70	0.0940	2.22	0.0339	2.74	0.0093
0.15	0.3945	0.67	0.3187	1.19	0.1965	1.71	0.0925	2.23	0.0332	2.75	0.0091
0.16	0.3939	0.68	0.3166	1.20	0.1942	1.72	0.0909	2.24	0.0325	2.76	0.0088
0.17	0.3932	0.69	0.3144	1.21	0.1919	1.73	0.0893	2.25	0.0317	2.77	0.0086
0.18	0.3925	0.70	0.3123	1.22	0.1895	1.74	0.0878	2.26	0.0310	2.78	0.0084
0.19	0.3918	0.71	0.3101	1.23	0.1872	1.75	0.0863	2.27	0.0303	2.79	0.0081
0.20	0.3910	0.72	0.3079	1.24	0.1849	1.76	0.0848	2.28	0.0297	2.80	0.0079
0.21	0.3902	0.73	0.3056	1.25	0.1826	1.77	0.0833	2.29	0.0290	2.81	0.0077
0.22	0.3894	0.74	0.3034	1.26	0.1804	1.78	0.0818	2.30	0.0283	2.82	0.0075
0.23	0.3885	0.75	0.3011	1.27	0.1781	1.79	0.0804	2.31	0.0277	2.83	0.0073
0.24	0.3876	0.76	0.2989	1.28	0.1758	1.80	0.0790	2.32	0.0270	2.84	0.0071
0.25	0.3867	0.77	0.2966	1.29	0.1736	1.81	0.0775	2.33	0.0264	2.85	0.0069
0.26	0.3857	0.78	0.2943	1.30	0.1714	1.82	0.0761	2.34	0.0258	2.86	0.0067
0.27	0.3847	0.79	0.2920	1.31	0.1691	1.83	0.0748	2.35	0.0252	2.87	0.0065
0.28	0.3836	0.80	0.2897	1.32	0.1669	1.84	0.0734	2.36	0.0246	2.88	0.0063
0.29	0.3825	0.81	0.2874	1.33	0.1647	1.85	0.0721	2.37	0.0241	2.89	0.0061
0.30	0.3814	0.82	0.2850	1.34	0.1626	1.86	0.0707	2.38	0.0235	2.90	0.0060
0.31	0.3802	0.83	0.2827	1.35	0.1604	1.87	0.0694	2.39	0.0229	2.91	0.0058
0.32	0.3790	0.84	0.2803	1.36	0.1582	1.88	0.0681	2.40	0.0224	2.92	0.0056
0.33	0.3778	0.85	0.2780	1.37	0.1561	1.89	0.0669	2.41	0.0219	2.93	0.0055
0.34	0.3765	0.86	0.2756	1.38	0.1539	1.90	0.0656	2.42	0.0213	2.94	0.0053
0.35	0.3752	0.87	0.2732	1.39	0.1518	1.91	0.0644	2.43	0.0208	2.95	0.0051
0.36	0.3739	0.88	0.2709	1.40	0.1497	1.92	0.0632	2.44	0.0203	2.96	0.0050
0.37	0.3725	0.89	0.2685	1.41	0.1476	1.93	0.0620	2.45	0.0198	2.97	0.0048
0.38	0.3712	0.90	0.2661	1.42	0.1456	1.94	0.0608	2.46	0.0194	2.98	0.0047
0.39	0.3697	0.91	0.2637	1.43	0.1435	1.95	0.0596	2.47	0.0189	2.99	0.0046
0.40	0.3683	0.92	0.2613	1.44	0.1415	1.96	0.0584	2.48	0.0184	3.00	0.0044
0.41	0.3668	0.93	0.2589	1.45	0.1394	1.97	0.0573	2.49	0.0180	3.10	0.0033
0.42	0.3653	0.94	0.2565	1.46	0.1374	1.98	0.0562	2.50	0.0175	3.20	0.0024
0.43	0.3637	0.95	0.2541	1.47	0.1354	1.99	0.0551	2.51	0.0171	3.30	0.0017
0.44	0.3621	0.96	0.2516	1.48	0.1334	2.00	0.0540	2.52	0.0167	3.40	0.0012
0.45	0.3605	0.97	0.2492	1.49	0.1315	2.01	0.0529	2.53	0.0163	3.50	0.0009
0.46	0.3589	0.98	0.2468	1.50	0.1295	2.02	0.0519	2.54	0.0158	3.60	0.0006
0.47	0.3572	0.99	0.2444	1.51	0.1276	2.03	0.0508	2.55	0.0154	3.70	0.0004
0.48	0.3555	1.00	0.2420	1.52	0.1257	2.04	0.0498	2.56	0.0151	3.80	0.0003
0.49	0.3538	1.01	0.2396	1.53	0.1238	2.05	0.0488	2.57	0.0147	3.90	0.0002
0.50	0.3521	1.02	0.2371	1.54	0.1219	2.06	0.0478	2.58	0.0143	4.00	0.0001
0.51	0.3503	1.03	0.2347	1.55	0.1200	2.07	0.0468	2.59	0.0139	4.10	0.0001

Use of the table: Having the sample mean $\overline{X}$ and empirical variance s_n^2 for the estimation of $E(X)$ and $D^2(X)$ of a Gaussian variable X, the construction of the density function can be carried out by using the formula

$$\varphi(y; \overline{X}, s_n) = \frac{1}{s_n} \varphi\left(\frac{y - \overline{X}}{s_n}\right).$$

This means that the value $\varphi(x)$ belonging to x is to be divided by s_n and the obtained value will be the ordinate to the abscissa $y = \overline{X} + xs_n$.

TABLE IV

DISTRIBUTION FUNCTION

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

OF THE STANDARD NORMAL VARIABLE

$\Phi(-x) = 1 - \Phi(x))$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.51	0.6950	1.02	0.8461	1.53	0.9370	2.08	0.9812
0.01	0.5040	0.52	0.6985	1.03	0.8485	1.54	0.9382	2.10	0.9821
0.02	0.5080	0.53	0.7019	1.04	0.8508	1.55	0.9394	2.12	0.9830
0.03	0.5120	0.54	0.7054	1.05	0.8531	1.56	0.9406	2.14	0.9838
0.04	0.5160	0.55	0.7088	1.06	0.8554	1.57	0.9418	2.16	0.9846
0.05	0.5199	0.56	0.7123	1.07	0.8577	1.58	0.9429	2.18	0.9854
0.06	0.5239	0.57	0.7157	1.08	0.8599	1.59	0.9441	2.20	0.9861
0.07	0.5279	0.58	0.7190	1.09	0.8621	1.60	0.9452	2.22	0.9868
0.08	0.5319	0.59	0.7224	1.10	0.8643	1.61	0.9463	2.24	0.9875
0.09	0.5359	0.60	0.7257	1.11	0.8665	1.62	0.9474	2.26	0.9881
0.10	0.5398	0.61	0.7291	1.12	0.8686	1.63	0.9484	2.28	0.9887
0.11	0.5438	0.62	0.7324	1.13	0.8708	1.64	0.9495	2.30	0.9893
0.12	0.5478	0.63	0.7352	1.14	0.8729	1.65	0.9505	2.32	0.9898
0.13	0.5517	0.64	0.7389	1.15	0.8749	1.66	0.9515	2.34	0.9904
0.14	0.5557	0.65	0.7422	1.16	0.8770	1.67	0.9525	2.36	0.9909
0.15	0.5596	0.66	0.7454	1.17	0.8790	1.68	0.9535	2.38	0.9913
0.16	0.5636	0.67	0.7486	1.18	0.8810	1.69	0.9545	2.40	0.9918
0.17	0.5675	0.68	0.7517	1.19	0.8830	1.70	0.9554	2.42	0.9922
0.18	0.5714	0.69	0.7549	1.20	0.8849	1.71	0.9564	2.44	0.9927
0.19	0.5753	0.70	0.7580	1.21	0.8869	1.72	0.9572	2.46	0.9931
0.20	0.5793	0.71	0.7611	1.22	0.8888	1.73	0.9582	2.48	0.9934
0.21	0.5832	0.72	0.7642	1.23	0.8907	1.74	0.9591	2.50	0.9938
0.22	0.5871	0.73	0.7673	1.24	0.8925	1.75	0.9599	2.52	0.9941
0.23	0.5910	0.74	0.7703	1.25	0.8944	1.76	0.9608	2.54	0.9945
0.24	0.5948	0.75	0.7734	1.26	0.8962	1.77	0.9616	2.56	0.9948
0.25	0.5987	0.76	0.7764	1.27	0.8980	1.78	0.9625	2.58	0.9951
0.26	0.6026	0.77	0.7794	1.28	0.8997	1.79	0.9633	2.60	0.9953
0.27	0.6064	0.78	0.7823	1.29	0.9015	1.80	0.9641	2.62	0.9956
0.28	0.6103	0.79	0.7853	1.30	0.9032	1.81	0.9649	2.64	0.9959
0.29	0.6141	0.80	0.7881	1.31	0.9049	1.82	0.9656	2.66	0.9961
0.30	0.6179	0.81	0.7910	1.32	0.9066	1.83	0.9664	2.68	0.9963
0.31	0.6217	0.82	0.7939	1.33	0.9082	1.84	0.9671	2.70	0.9965
0.32	0.6255	0.83	0.7967	1.34	0.9099	1.85	0.9678	2.72	0.9967
0.33	0.6293	0.84	0.7995	1.35	0.9115	1.86	0.9686	2.74	0.9969
0.34	0.6331	0.85	0.8023	1.36	0.9131	1.87	0.9693	2.76	0.9971
0.35	0.6368	0.86	0.8051	1.37	0.9147	1.88	0.9699	2.78	0.9973
0.36	0.6406	0.87	0.8078	1.38	0.9162	1.89	0.9706	2.80	0.9974
0.37	0.6443	0.88	0.8106	1.39	0.9177	1.90	0.9713	2.82	0.9976
0.38	0.6480	0.89	0.8133	1.40	0.9192	1.91	0.9719	2.84	0.9977
0.39	0.6517	0.90	0.8159	1.41	0.9207	1.92	0.9726	2.86	0.9979
0.40	0.6554	0.91	0.8186	1.42	0.9222	1.93	0.9732	2.88	0.9980
0.41	0.6591	0.92	0.8212	1.43	0.9236	1.94	0.9738	2.90	0.9981
0.42	0.6628	0.93	0.8238	1.44	0.9251	1.95	0.9744	2.92	0.9982
0.43	0.6664	0.94	0.8264	1.45	0.9265	1.96	0.9750	2.94	0.9984
0.44	0.6700	0.95	0.8289	1.46	0.9279	1.97	0.9756	2.96	0.9985
0.45	0.6736	0.96	0.8315	1.47	0.9292	1.98	0.9761	2.98	0.9986
0.46	0.6772	0.97	0.8340	1.48	0.9306	1.99	0.9767	3.00	0.9987
0.47	0.6808	0.98	0.8365	1.49	0.9319	2.00	0.9772	3.20	0.9993
0.48	0.6844	0.99	0.8389	1.50	0.9332	2.02	0.9783	3.40	0.9996
0.49	0.6879	1.00	0.8413	1.51	0.9345	2.04	0.9793	3.60	0.9998
0.50	0.6915	1.01	0.8438	1.52	0.9357	2.06	0.9803	3.80	0.9999

Hint: The distribution function of $N(\mu, \sigma)$ can be obtained from the formula

$$\Phi(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y-\mu}{\sigma}} e^{-\frac{t^2}{2}}\, dt.$$

This means that the value of the distribution function belonging to y is $\Phi\left(\frac{y-\mu}{\sigma}\right)$.

TABLE Va

PROBABILITY $p_{\lambda}(k)$ OF THE VALUE k OF A POISSON-DISTRIBUTED VARIABLE

k \ λ	0.1	0.2	0.3	0.4	0.5
0	0.90484	0.81873	0.74082	0.67032	0.60653
1	0.09048	0.16375	0.22225	0.26813	0.30327
2	0.00452	0.01637	0.03334	0.05362	0.07581
3	0.00015	0.00109	0.00333	0.00715	0.01263
4		0.00005	0.00025	0.00071	0.00158
5			0.00001	0.00005	0.00016
6					0.00001

k \ λ	0.6	0.7	0.8	0.9
0	0.54881	0.49659	0.44933	0.40657
1	0.32929	0.34761	0.35946	0.36591
2	0.09878	0.12166	0.14379	0.16466
3	0.01976	0.02838	0.03834	0.04939
4	0.00296	0.00496	0.00766	0.01111
5	0.00035	0.00069	0.00123	0.00200
6	0.00003	0.00008	0.00016	0.00030
7			0.00001	0.00003

k \ λ	1	2	3	4	5
0	0.36788	0.13534	0.04978	0.01831	0.00673
1	0.36788	0.27067	0.14936	0.07326	0.03369
2	0.18394	0.27067	0.22404	0.14653	0.08422
3	0.06131	0.18045	0.22404	0.19537	0.14037
4	0.01532	0.09022	0.16803	0.19537	0.17547
5	0.00306	0.03609	0.10082	0.15629	0.17547
6	0.00051	0.01203	0.05040	0.10420	0.14622
7	0.00007	0.00343	0.02160	0.05954	0.10444
8		0.00085	0.00810	0.02977	0.06527
9		0.00019	0.00270	0.01322	0.03626
10		0.00003	0.00081	0.00529	0.01813
11			0.00022	0.00192	0.00824
12			0.00005	0.00064	0.00343
13			0.00001	0.00019	0.00132
14				0.00005	0.00047
15				0.00001	0.00015
16					0.00004
17					0.00001

k \ λ	6	7	8	9	10
0	0.00247	0.00091	0.00033	0.00012	0.00004
1	0.01487	0.00638	0.00268	0.00111	0.00045
2	0.04461	0.02234	0.01073	0.00499	0.00227
3	0.08923	0.05212	0.02862	0.01499	0.00756
4	0.13385	0.09122	0.05725	0.03373	0.01891
5	0.16062	0.12772	0.09160	0.06072	0.03783
6	0.16062	0.14900	0.12214	0.09109	0.06305
7	0.13768	0.14900	0.13959	0.11712	0.09007
8	0.10326	0.13038	0.13959	0.13176	0.11260
9	0.06883	0.10140	0.12408	0.13176	0.12511
10	0.04130	0.07098	0.09926	0.11858	0.15511
11	0.02252	0.04517	0.07219	0.09702	0.11374
12	0.01126	0.02635	0.04812	0.07276	0.09478
13	0.00519	0.01418	0.02961	0.05037	0.07290
14	0.00222	0.00709	0.01692	0.03238	0.05207
15	0.00089	0.00331	0.00902	0.01943	0.03471
16	0.00033	0.00144	0.00451	0.01093	0.02169
17	0.00011	0.00059	0.00212	0.00578	0.01276
18	0.00003	0.00023	0.00094	0.00289	0.00709
19	0.00001	0.00008	0.00039	0.00137	0.00373
20		0.00003	0.00015	0.00061	0.00183
21			0.00006	0.00026	0.00088
22			0.00002	0.00010	0.00040
23				0.00004	0.00017
24				0.00001	0.00007
25					0.00002
26					0.00001

TABLE Vb

CUMULATIVE SUMS (DISTRIBUTION) OF THE POISSON PROBABILITIES

k \ λ	0.1	0.2	0.3	0.4	0.5	0.6
0	0.904837	0.818731	0.740818	0.670320	0 606531	0.548812
1	0.995321	0.982477	0.963063	0.938448	0.909796	0.878099
2	0.999845	0.998852	0.996390	0.992074	0.985612	0.977885
3	0.999996	0.999943	0.999724	0.999224	0.998248	0.997642
4	1.000000	0.999998	0.999974	0.999939	0.999828	0.999606
5	1.000000	1.000000	0.999999	0.999996	0.999986	0.999962
6	1.000000	1.000000	1.000000	1.000000	0.999999	0.999997
7	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

k \ λ	0.7	0.8	0.9	1.0	2.0	3.0
0	0.496585	0.449329	0.406570	0.367879	0.135335	0.049787
1	0.844195	0.808792	0.772483	0.735759	0.406006	0.199148
2	0.965858	0.952577	0.937144	0.919699	0.676677	0.423190
3	0.994246	0.990920	0.988542	0.981012	0.857124	0.647232
4	0.999214	0.998589	0.997657	0.996340	0.947348	0.815263
5	0.999909	0.999816	0.999658	0.909406	0.983437	0.916082
6	0.999990	0.999980	0.999958	0.999917	0.995467	0.966491
7	0.999998	0.999999	0.999997	0.999990	0.998904	0.988095
8	1.000000	1.000000	1.000000	0.999999	0.999763	0.996196
9				1.000000	0.999954	0.998897
10					0.999992	0.999707
11					0.999999	0.999928
12					1.000000	0.999998
13						0.999996
14						1.000009

k \ λ	4.0	5.0	6.0	7.0	8.0	9.0
0	0.018316	0.006738	0.002479	0.000912	0.000335	0.000123
1	0.091579	0.040428	0.017352	0.007295	0.003019	0.001234
2	0.238105	0.124652	0.061970	0.029636	0.013754	0.006232
3	0.433472	0.265026	0.151205	0.081765	0.042380	0.021228
4	0.628839	0.440493	0.285058	0.172991	0.099632	0.054963
5	0.785132	0.615960	0.445681	0.300708	0.191236	0.115690
6	0.889326	0.762183	0.606304	0.449711	0.313374	0.206780
7	0.948866	0.866628	0.743981	0.598714	0.452961	0.323896
8	0.978636	0.931806	0.847239	0.729091	0.592548	0.455652
9	0.991867	0.968172	0.916077	0.830496	0.716625	0.587408
10	0.997159	0.986205	0.957380	0 901479	0.815887	0.705988
11	0.999084	0.994547	0.979909	0.946650	0.888077	0.803008
12	0.999726	0.997981	0.991173	0.973000	0.936204	0.875773
13	0.999923	0.999202	0.996372	0.987188	0.965820	0.926149
14	0.999979	0.999774	0.998600	0.994282	0.982744	0.958533
15	0.999994	0.999931	0.999491	0.997593	0.991770	0.977964
16	0.999998	0.999980	0.999825	0.999041	0.996283	0.988894
17	0.999999	0.999994	0.999943	0.999637	0.998407	0.994680
18	0.999999	0.999998	0.999982	0.999869	0.999351	0.997573
19	0.999999	0.999999	0.999994	0.999955	0.999748	0.998943
20	1.000000	0.999999	0.999998	0.999985	0.999907	0.999560
21		1.000000	0.999999	0.999995	0.999967	0.999824
22			0.999999	0.999998	0.999989	0.999932
23			1.000000	0.999999	0.999997	0.999974
24				0.999999	.0999999	0.999990
25				1.000000	0.999999	0.999996
26					1.000000	0.999998
27						0.999999
28						1.000000

TABLE VIa

PROBABILITY $p_n(k)$ OF THE VALUE k IN THE CASE OF THE (n, p) BINOMIAL DISTRIBUTION

n	k	p											
		0.01	0.02	0.04	0.05	0.08	0.1	0.2	0.3	0.4	0.5		
5	0	0.95099	0.90392	0.81537	0.77378	0.65908	0.59049	0.32768	0.16807	0.07776	0.03125	5	5
	1	0.04803	0-09224	0.16988	0.20363	0.28656	0.32805	0,40960	0)36015	0.65920	0.25625	4	
	2	0.00097	0.00376	0.01416	0.02143	0.04984	0.07290	0.20480	0.30870	0.34560	0.31250	3	
	3	0 00001	0.00008	0.00059	0 00113	0 00433	0 00810	0 05120	0 13230	0.23040	0.31250	2	
	4			0 00001	0 00003	0 00019	0 00045	0 00640	0 02835	0 07680	0.15625	1	
	5						0 00001	0 00032	0 00243	0 01024	0 03125	0	
10	0	0.90438	0.81707	0.66483	0.59874	0.43439	0.34868	0.10737	0.02825	0.00605	0.00098	10	10
	1	0.09135	0.16675	0.27701	0.31512	0.37773	0.38742	0.26844	0.12106	0.04041	0.00977	9	
	2	0.00415	0.01531	0.05194	0.07463	0.14781	0.19371	0.30199	0.23357	0.12093	0.04395	8	
	3	0.00011	0.00083	0.00577	0.01048	0.03427	0.05740	0.20133	0.26683	0.21499	0.11719	7	
	4		0.00003	0.00042	0.00096	0.00522	0.01116	0.08808	0.20012	0.25082	0.20508	6	
	5			0.00002	0.00006	0.00054	0.00149	0.02642	0.10292	0.20066	0.24609	5	
	6					0.00004	0.00014	0.00551	0.03676	0.11148	0.20508	4	
	7						0.00001	0.00079	0.00900	0.04247	0.11719	3	
	8							0.00007	0.00145	0.01062	0.04395	2	
	9								0.00014	0.00157	0.00977	1	
	10								0.00001	0.00010	0.00098	0	
15	0	0.86006	0.73857	0.54209	0.46329	0.28630	0.20589	0.03518	0.00475	0.00047	0.00003	15	15
	1	0.13031	0.22609	0.33880	0.36576	0.37343	0.34315	0.13194	0.03052	0.00470	0.00046	14	
	2	0.00921	0.03230	0.09882	0.13475	0.22731	0.26690	0.23090	0.09156	0.02194	0.00320	13	
	3	0.00040	0.00286	0.01784	0.03073	0.08565	0.12851	0.25014	0.17004	0.06339	0.01389	12	
	4	0.00001	0.00017	0.00223	0.00485	0.02234	0.04284	0.18760	0.21862	0.12678	0.04166	11	
	5		0.00001	0.00020	0.00056	0.00427	0.01047	0.10318	0.20613	0.18594	0.09164	10	
	6			0.00001	0.00005	0.00062	0.00194	0.04299	0.14724	0.20660	0.15274	9	
	7					0.00007	0.00028	0.01382	0.08113	0.17708	0.19638	8	
	8					0.00001	0.00003	0.00345	0.03477	0.11806	0.19638	7	
	9							0.00067	0.01159	0.06121	0.15274	6	
	10							0.00010	0.00298	0.02449	0.09164	5	
	11							0.00001	0.00058	0.00742	0.04166	4	
	12								0.00008	0.00165	0.01389	3	
	13								0.00001	0.00025	0.00320	2	
	14									0.00002	0.00046	1	
	15										0.00003	0	
20	0	0.81791	0.66761	0.44200	0.35849	0.18869	0.12158	0.01153	0.00080	0.00004		20	20
	1	0.16523	0.27249	0.36834	0.37735	0.32816	0.27017	0.05765	0.00684	0.00049	0.00002	19	
	2	0.01586	0.05283	0.14580	0.18868	0.27109	0.28518	0.13691	0.02785	0.00309	0.00018	18	
	3	0.00096	0.00647	0.03645	0.05958	0.14144	0.19012	0.20536	0.07160	0.01235	0.00109	17	
	4	0.00004	0.00056	0.00645	0.01333	0.05227	0.08978	0.21820	0.13042	0.03499	0.00462	16	
	5		0.00004	0.00086	0.00224	0.01454	0.03192	0.17456	0.17886	0.07465	0.01479	15	
	6			0.00009	0.00030	0.00316	0.00887	0.10910	0.19164	0.12441	0.03696	14	
	7			0.00001	0.00003	0.00055	0.00197	0.05455	0.16426	0.16588	0.07393	13	
	8					0.00008	0.00036	0.02216	0.11440	0.17971	0.12013	12	
	9					0.00001	0.00005	0.00739	0.06537	0.15974	0.16018	11	
	10						0.00001	0.00203	0.03082	0.11714	0.17620	10	
	11							0.00046	0.01201	0.07099	0.16018	9	
	12							0.00009	0.00386	0.03550	0.12013	8	
	13							0.00001	0.00102	0.01456	0.07393	7	
	14								0.00022	0.00485	0.03696	6	
	15								0.00004	0.00129	0.01479	5	
	16								0.00001	0.00027	0.00462	4	
	17									0.00004	0.00109	3	
	18										0.00018	2	
	20										0.00002	1	
												0	
		0.99	0.98	0.96	0.95	0.92	0.9	0.8	0.7	0.6	0.5	k	n
		p											

TABLE VIa (cont.)

n	k	p											
		0.01	0.02	0.04	0.05	0.08	0.1	0.2	0.3	0.4	0.5		
25	0	0.77782	0.60346	0.36040	0.27739	0 12436	0.07179	0.00378	0.00013			25	25
	1	0.19642	0.30789	0.37541	0.36499	0.27036	0.19942	0.02361	0.00144	0.00005		24	
	2	0.02381	0.07540	0.18771	0.23052	0.28211	0.26589	0.07084	0.00739	0.00038	0.00001	23	
	3	0.00184	0.01180	0.05996	0.09302	0.18807	0.22650	0.13577	0.02428	0.00194	0.00007	22	
	4	0.00010	0.00132	0.01374	0.02693	0.08995	0.13842	0.18668	0.05723	0.00710	0.00038	21	
	5	0.00000	0.00011	0.00240	0.00595	0.03285	0.06459	0.19602	0.10302	0.01989	0.00158	20	
	6		0.00001	0.00033	0.00104	0.00952	0.02392	0.16335	0.14717	0.04420	0.00528	19	
	7			0.00004	0.00015	0.00225	0.00722	0.11084	0.17119	0.07999	0.01433	18	
	8				0.00002	0.00044	0.00180	0.06235	0.16508	0.11998	0.03223	17	
	9					0.00007	0.00038	0.02944	0.13364	0.15109	0.06089	16	
	10					0.00001	0.00097	0.01178	0.09164	0.16116	0.09742	15	
	11						0.00001	0.00401	0.05355	0.14651	0.13284	14	
	12							0.00117	0.02678	0.11395	0.15498	13	
	13							0.00029	0.01148	0.07597	0.15498	12	
	14							0.00006	0.00422	0.04341	0.13284	11	
	15							0.00001	0.00132	0.02122	0.09742	10	
	16								0.00035	0.00884	0.06089	9	
	17								0.00008	0.00312	0.03223	8	
	18								0.00002	0 00092	0.01433	7	
	19									0.00023	0.00528	6	
	20									0.00005	0.00158	5	
	21									0.00001	0.00038	4	
	22										0.00007	3	
30	0	0 73970	0,54548	0,29386	0,21464	0,08197	0,04239	0,00124	0,00002			30	30
	1	0.22415	0.33397	0.36732	0.33890	0.21382	0.14130	0.00928	0.00029			29	
	2	0.03283	0.09883	0.22192	0.25864	0.26961	0.22766	0.03360	0.00180	0.00004		28	
	3	0.00310	0.01882	0.08630	0.12705	0.21881	0.23609	0.07853	0.00720	0.00027		27	
	4	0.00021	0.00259	0.02427	0.04514	0.12843	0.17707	0.13252	0.02084	0.00120	0.00003	26	
	5	0.00001	0.00028	0.00526	0.01235	0.05807	0.10230	0.17228	0.04644	0.00415	0.00013	25	
	6		0.00002	0.00091	0.00271	0.02104	0.04736	0.17946	0.08293	0.01152	0.00055	24	
	7			0.00013	0.00049	0.00627	0.01804	0.15382	0.12185	0.02634	0.00190	23	
	8			0.00002	0.00007	0.00157	0.00576	0.11056	0.15014	0.05049	0.00545	22	
	9				0.00001	0.00033	0.00157	0.06756	0.15729	0.08228	0.01332	21	
	10					0.00006	0.00037	0.03547	0.14156	0.11519	0.02798	20	
	11					0.00001	0.00007	0.01612	0.11031	0.13962	0.05088	19	
	12						0.00001	.00638	0.07485	0.14738	0.08055	18	
	13							0.00221	0.04442	0.13604	0.11154	17	
	14							0.00067	0.0231	0.11013	0.13544	16	
	15							0.00018	0.01057	0.07831	0.14446	15	
	16							0.00004	0.00425	0.04895	0.13544	14	
	17							0.00001	0.00170	0.02687	0.11154	13	
	18								0.00046	0.01294	0.08055	12	
	19								0.00013	0.00545	0.05088	11	
	20								0.00003	0.00200	0.02798	10	
	21								0.00001	0.00063	0.04332	9	
	22									0.00017	0.00545	8	
	23									0,00004	0.00190	7	
	24									0.00001	0.00055	6	
	25										0.00013	5	
	26										0.00003	4	
		0.99	0.98	0.96	0.95	0.92	0.9	0.8	0.8	0.7	0.5		
		p											

TABLE VIb

DISTRIBUTION FUNCTION OF THE BINOMIAL LAW

n	k	p										k	n
		0.01	0.02	0.04	0.05	0.08	0.1	0.2	0.3	0.4	0.5		
5	0	0.95099	0.90392	0.81537	0.77378	0.65908	0.59409	0.32768	0.16807	0.07776	0.03125	0	5
	1	0.99902	0.99616	0.98524	0.97741	0.94564	0.91854	0.73728	0.52822	0.33696	0.18750	1	
	2	0.99999	0.99992	0.99940	0.99884	0.99548	0.99144	0.94208	0.83692	0.68256	0.50000	2	
	3	1.00000	1.00000	0.99999	0.99997	0.99981	0.99954	0.99328	0.96922	0.91296	0.81250	3	
	4			1.00000	1.00000	1.00000	0.99999	0.99968	0.99757	0.98976	0.96875	4	
	5						1.00000	1.00000	1.00000	1.00000	1.00000	5	
10	0	0.90438	0.81707	0.66483	0.59874	0.43439	0.34868	0.10737	0.02825	0.00605	0.00098	0	10
	1	0.99574	0.98382	0.94184	0.91386	0.81212	0.73610	0.37581	0.14931	0.04636	0.01075	1	
	2	0.99989	0.99914	0.99378	0.98850	0.95993	0.92981	0.67780	0.38278	0.16729	0.05470	2	
	3	1.00000	0.99997	0.99956	0.99897	0.99420	0.98721	0.87913	0.64961	0.38228	0.17189	3	
	4		1.00000	0.99998	0.99994	0.99942	0.98836	0.96721	0.84973	0.63310	0.37697	4	
	5			1.00000	1.00000	0.99996	0.99985	0.99363	0.95265	0.83376	0.62306	5	
	6					1.00000	0.99999	0.99914	0.98940	0.94524	0.82814	6	
	7						1.00000	0.99993	0.99840	0.98771	0.94533	7	
	8							1.00000	0.99985	0.99833	0.98928	8	
	9								0.99999	0.99990	0.99905	9	
	10								1.00000	1.00000	1.00000	10	
15	0	0.86006	0.73857	0.54209	0.46329	0.28630	0.20589	0.03518	0.00475	0.00047	0.00003	0	15
	1	0.99037	0.96466	0.88089	0.82905	0.65973	0.54904	0.16712	0.03527	0.00517	0.00049	1	
	2	0.99958	0.99696	0.97971	0.96380	0.88704	0.81594	0.39802	0.12683	0.02711	0.00369	2	
	3	0.99999	0.99982	0.99755	0.99453	0.97269	0.94445	0.64816	0.29687	0.09050	0.01758	3	
	4	1.00000	0.99999	0.99979	0.99939	0.99503	0.98729	0.83576	0.51549	0.21728	0.05924	4	
	5		1.00000	0.99999	0.99995	0.99930	0.99776	0.93894	0.72162	0.40322	0.15088	5	
	6			1.00000	1.00000	0.99992	0.99969	0.98193	0.86886	0.60982	0.30362	6	
	7					0.99999	0.99997	0.99575	0.94999	0.78690	0.50000	7	
	8					1.00000	1.00000	0.99920	0.98476	0.90496	0.69638	8	
	9							0.99987	0.99635	0.96617	0.84912	9	
	10							0.99999	0.99933	0.99066	0.94076	10	
	11							1.00000	0.99991	0.99808	0.98242	11	
	12								0.99999	0.99973	0.99631	12	
	13								1.00000	0.99998	0.99951	13	
	14									1.00000	0.99997	14	
	15										1.00000	15	
20	0	0.81791	0.66761	0.44200	0.35849	0.18869	0.12158	0.01153	0.00080	0.00004		0	20
	1	0.98314	0.94010	0.81034	0.73584	0.51685	0.39175	0.06918	0.00764	0.00053	0.00002	1	
	2	0.99900	0.99293	0.95614	0.92452	0.78794	0.67693	0.20609	0.03549	0.00362	0.00020	2	
	3	0.99996	0.99940	0.99259	0.98410	0.92938	0.86705	0.41145	0.10709	0.01597	0.00129	3	
	4	1.00000	0.99996	0.99904	0.99743	0.98165	0.95683	0.62965	0.23751	0.05096	0.00591	4	
	5		1.00000	0.99990	0.99967	0.99619	0.98875	0.80421	0.41637	0.12561	0.02070	5	
	6			0.99999	0.99997	0.99935	0.99761	0.91331	0.60801	0.25002	0.05766	6	
	7			1.00000	1.00000	0.99991	0.99958	0.96786	0.77227	0.41590	0.13159	7	
	8					0.99999	0.99994	0.99002	0.88667	0.59561	0.25172	8	
	9					1.00000	0.99999	0.99741	0.95204	0.75535	0.41190	9	
	10						1.00000	0.99944	0.98286	0.87249	0.58810	10	
	11							0.99990	0.99487	0.94348	0.74828	11	
	12							0.99999	0.99872	0.97898	0.86841	12	
	13							1.00000	0.99973	0.99354	0.94234	13	
	14								0.99995	0.99840	0.97930	14	
	15								0.99999	0.99969	0.99409	15	
	16								1.00000	0.99996	0.99871	16	
	17									1.00000	0.99980	17	
	18										0.99998	19	
	19										1.00000	18	
	20											20	

TABLE VIb (cont.)

n	k	p										k	n
		0.01	0.02	0.04	0.05	0.08	0.1	0.2	0.3	0.4	0.5		
25	0	0.77782	0.60346	0.36040	0.27739	0.12436	0.07179	0.00378	0.00013			0	25
	1	0.97424	0.91135	0.73581	0.64238	0.39472	0.27121	0.02739	0.00157	0.00005		1	
	2	0.99805	0.98675	0.92352	0.87289	0.67683	0.53710	0.09823	0.00896	0.00043	0.00001	2	
	3	0.99989	0.99856	0.98348	0.96591	0.86490	0.76360	0.23400	0.03324	0.00237	0.00008	3	
	4	0.99999	0.99988	0.99723	0.99283	0.95485	0.90202	0.42068	0.09047	0.00946	0.00046	4	
	5	1.00000	0.99999	0.99963	0.99878	0.98770	0.96661	0.61670	0.19349	0.02936	0.00204	5	
	6		1.00000	0.99996	0.99982	0.99722	0.99053	0.78005	0.34066	0.07356	0.00732	5	
	7			1.00000	0.99998	0.99948	0.99775	0.89089	0.51185	0.15355	0.02165	7	
	8				1.00000	0.99992	0.99954	0.95324	0.67693	0.27353	0.05388	8	
	9					0.99999	0.99992	0.98268	0.81057	0.42462	0.11477	9	
	10					1.00000	0.99999	0.99446	0.90221	0.58578	0.21219	10	
	11						1.00000	0.99847	0.95576	0.73229	0.34503	11	
	12							0.99964	0.98254	0.84624	0.50001	12	
	13							0.99993	0.99402	0.92221	0.65499	13	
	14							0.99999	0.99824	0.96562	0.78783	14	
	15							1.00000	0.99955	0.98684	0.88525	15	
	16								0.99990	0.99568	0.94612	16	
	17								0.99998	0.99880	0.97838	17	
	18								1.00000	0.99972	0.99268	18	
	19									0.99994	0.99796	19	
	20									0.99999	0.99954	20	
	21									1.00000	0.99992	21	
	22										0.99999	22	
	23										1.00000	23	
30	0	0.73970	0.54548	0.29386	0.21464	0.08197	0.04239	0.00124	0.00002			0	30
	1	0.96385	0.87945	0.66118	0.55354	0.29579	0.18369	0.01052	0.00031			1	
	2	0.99668	0.97828	0.88310	0.81218	0.56540	0.41135	0.04418	0.00211	0.00004		2	
	3	0.99978	0.99710	0.96940	0.93923	0.78421	0.64744	0.12271	0.00931	0.00031		3	
	4	0.99999	0.99970	0.99367	0.98436	0.91264	0.82451	0.25523	0.03015	0.00151	0.00003	4	
	5	1.00000	0.99998	0.99893	0.99672	0.97071	0.92681	0.42751	0.07659	0.00566	0.00016	5	
	6		1.00000	0.99985	0.99943	0.99175	0.97417	0.60697	0.15952	0.01718	0.00071	6	
	7			0.99998	0.99992	0.99802	0.99221	0.76079	0.28137	0.04352	0.00261	7	
	8			1.00000	0.99999	0.99960	0.99797	0.87135	0.43151	0.09401	0.00806	8	
	9				1.00000	0.99993	0.99955	0.93891	0.58880	0.17629	0.02138	9	
	10					0.99999	0.99992	0.97438	0.73036	0.29148	0.04936	10	
	11					1.00000	0.99999	0.99050	0.84067	0.43110	0.10024	11	
	12						1.00000	0.99688	0.91552	0.57848	0.18079	12	
	13							0.99909	0.95994	0.71452	0.29233	13	
	14							0.99977	0.98306	0.82465	0.42777	14	
	15							0.99995	0.99363	0.90296	0.57223	15	
	16							0.99999	0.99788	0.95191	0.70767	16	
	17							1.00000	0.99938	0.97878	0.81921	17	
	18								0.99983	0.99171	0.89976	18	
	19								0.99996	0.99716	0.95064	19	
	20								0.99999	0.99915	0.97862	20	
	21								1.00000	0.99978	0.99194	21	
	22									0.99995	0.99739	22	
	23									0.99999	0.99929	23	
	24									1.00000	0.99984	24	
	25										0.99997	25	
	26										1.00000	26	

TABLE

TABLE FOR THE *F* TEST AT ONE-(TWO-)SIDED

Critical values of the *F* test at one-sided

Parameter f_2 of the denominator	Parameter (degree of freedom) f_1 of the numerator																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246	246	247	247
2	18.5	19.0	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	18.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67
4	7.61	6.94	6.95	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17
9	5.12	4.26	3.86	3.63	3,48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2,97	2.96
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.20
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.26
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.12
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.18
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.23	2.20	2.18	2.15	2.13	2.11	2.09	2.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18	2.15	2.13	2.11	2.09	2.07	2.05
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05	2.02	1.99	1.97	1.95	1.93	1.92
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03	2.00	1.98	1.95	1.93	1.92	1.90
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.99	1.96	1.94	1.92	1.90	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99	1.96	1.93	1.91	1.89	1.87	1.86
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.95	1.92	1.90	1.88	1.86	1.84
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97	1.94	1.91	1.89	1.87	1.85	1.83
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96	1.93	1.90	1.88	1.86	1.84	1.82
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87	1.85	1.83	1.81
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93	1.90	1.88	1.85	1.83	1.81	1.79
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84	1.82	1.80	1.78
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84	1.81	1.79	1.77	1.75
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82	1.79	1.77	1.75	1.73
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83	1.80	1.78	1.76	1.74	1.72
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77	1.75	1.73	1.71
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.87	1.83	1.80	1.77	1.75	1.72	1.70	1.69
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.79	1.76	1.73	1.71	1.69	1.67
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74	1.72	1.69	1.67	1.66
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.82	1.78	1.75	1.72	1.70	1.68	1.66	1.64
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.77	1.74	1.71	1.69	1.66	1.64	1.62
1000	3.85	3.00	2.62	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.73	1.70	1.68	1.65	1.63	1.61
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69	1.67	1.64	1.62	1.60

E.g. If the parameters are $f_1 = 8$ and $f_2 = 20$ then $P(F \leqslant 2.45) = 95\%$.

VII

LEVELS 95% (90%), 99% (98%), 99.9% (99.8%)

level 95%, or two-sided level 90%

Parameter (degree of freedom) f_1 of the numerator

19	20	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞	Parameter f_2 of the denominator
248	248	249	249	249	250	250	251	251	251	252	252	252	253	254	254	254	1
19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	2
8.67	8.66	8.65	8.64	8.63	8.62	8.62	8.60	8.59	8.59	8.58	8.57	8.56	8.55	8.54	8.53	8.53	3
5.81	5.80	5.79	5.77	5.76	5.75	5.75	5.73	5.72	5.71	5.70	5.69	5.67	5.66	5.65	5.64	5.63	4
4.57	4.56	4.54	4.53	4.52	4.50	4.50	4.48	4.46	4.45	4.44	4.43	4.41	4.41	4.39	4.37	4.37	5
3.88	3.87	3.86	3.84	3.83	3.82	3.81	3.79	3.77	3.76	3.75	3.74	3.72	3.71	3.69	3.68	3.67	6
3.46	3.44	3.43	3.41	3.40	3.39	3.38	3.36	3.34	3.33	3.32	3.30	3.29	3.27	3.25	3.24	3.23	7
3.16	3.15	3.13	3.12	3.10	3.09	3.08	3.06	3.04	3.03	3.02	3.01	2.99	2.97	2.95	2.94	2.93	8
2.95	2.94	2.92	2.90	2.89	2.87	2.86	2.84	2.83	2.81	2.80	2.79	2.77	2.76	2.73	2.72	2.71	9
2.78	2.77	2.75	2.74	2.72	2.71	2.70	2.68	2.66	2.65	2.64	2.62	2.60	2.59	2.56	2.55	2.54	10
2.66	2.65	2.63	2.61	2.59	2.58	2.57	2.55	2.53	2.52	2.51	2.49	2.47	2.46	2.43	2.42	2.40	11
2.56	2.54	2.52	2.51	2.49	2.48	2.47	2.44	2.43	2.41	2.40	2.38	2.36	2.35	2.32	2.31	2.30	12
2.47	2.46	2.44	2.42	2.41	2.39	2.38	2.36	2.34	2.33	2.31	2.30	2.27	2.26	2.23	2.22	2.21	13
2.40	2.39	2.37	2.35	2.33	2.32	2.31	2.28	2.27	2.25	2.24	2.22	2.20	2.19	2.16	2.14	2.13	14
2.34	2.33	2.31	2.29	2.27	2.26	2.25	2.22	2.20	2.19	2.18	2.16	2.14	2.12	2.10	2.08	2.07	15
2.29	2.28	2.25	2.24	2.22	2.21	2.19	2.17	2.15	2.14	2.12	2.11	2.08	2.07	2.04	2.02	2.01	16
2.24	2.23	2.21	2.19	2.17	2.16	2.15	2.12	2.10	2.09	2.08	2.06	2.03	2.02	1.99	1.97	1.96	17
2.20	2.19	2.17	2.15	2.13	2.12	2.11	2.08	2.06	2.05	2.04	2.02	1.99	1.98	1.95	1.93	1.92	18
2.17	2.16	2.13	2.11	2.10	2.08	2.07	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.91	1.89	1.88	19
2.14	2.12	2.10	2.08	2.07	2.05	2.04	2.01	1.99	1.98	1.97	1.95	1.92	1.91	1.88	1.86	1.84	20
2.11	2.10	2.07	2.05	2.04	2.02	2.01	1.98	1.96	1.95	1.94	1.92	1.89	1.88	1.84	1.82	1.81	21
2.08	2.07	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.92	1.91	1.89	1.86	1.85	1.82	1.80	1.78	22
2.06	2.05	2.02	2.00	1.99	1.97	1.96	1.93	1.91	1.90	1.88	1.86	1.84	1.82	1.79	1.77	1.76	23
2.04	2.03	2.00	1.98	1.97	1.95	1.94	1.91	1.89	1.88	1.86	1.84	1.82	1.80	1.77	1.75	1.73	24
2.02	2.01	1.98	1.96	1.95	1.93	1.92	1.89	1.87	1.86	1.84	1.82	1.80	1.78	1.75	1.73	1.71	25
2.00	1.99	1.97	1.95	1.93	1.91	1.90	1.87	1.85	1.84	1.82	1.80	1.78	1.76	1.73	1.71	1.69	26
1.99	1.97	1.95	1.93	1.91	1.90	1.88	1.86	1.84	1.82	1.81	1.79	1.76	1.74	1.71	1.69	1.67	27
1.97	1.96	1.93	1.91	1.90	1.88	1.87	1.84	1.82	1.80	1.79	1.77	1.74	1.73	1.69	1.67	1.65	28
1.96	1.94	1.92	1.90	1.88	1.87	1.85	1.83	1.81	1.79	1.77	1.75	1.73	1.71	1.67	1.65	1.64	29
1.95	1.93	1.91	1.89	1.87	1.85	1.84	1.81	1.79	1.77	1.76	1.74	1.71	1.70	1.66	1.64	1.62	30
1.92	1.91	1.88	1.86	1.85	1.83	1.82	1.79	1.77	1.75	1.74	1.71	1.69	1.67	1.63	1.61	1.59	32
1.90	1.89	1.86	1.84	1.82	1.80	1.80	1.77	1.75	1.73	1.71	1.69	1.66	1.65	1.61	1.59	1.57	34
1.88	1.87	1.85	1.82	1.81	1.79	1.78	1.75	1.73	1.71	1.69	1.67	1.64	1.62	1.59	1.56	1.55	36
1.87	1.85	1.83	1.81	1.79	1.77	1.76	1.73	1.71	1.69	1.68	1.65	1.62	1.61	1.57	1.54	1.53	38
1.85	1.84	1.81	1.79	1.77	1.76	1.74	1.72	1.69	1.67	1.66	1.64	1.61	1.59	1.55	1.53	1.51	40
1.84	1.83	1.80	1.78	1.76	1.74	1.73	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.53	1.51	1.49	42
1.83	1.81	1.79	1.77	1.75	1.73	1.72	1.69	1.67	1.65	1.63	1.61	1.58	1.56	1.52	1.49	1.48	44
1.82	1.80	1.78	1.76	1.74	1.72	1.71	1.68	1.65	1.64	1.62	1.60	1.57	1.55	1.51	1.48	1.46	46
1.81	1.79	1.77	1.75	1.73	1.71	1.70	1.67	1.64	1.62	1.61	1.59	1.56	1.54	1.49	1.47	1.45	48
1.80	1.78	1.76	1.74	1.72	1.70	1.69	1.66	1.63	1.61	1.60	1.58	1.54	1.52	1.48	1.46	1.44	50
1.78	1.76	1.74	1.72	1.70	1.68	1.67	1.64	1.61	1.59	1.58	1.55	1.52	1.50	1.46	1.43	1.41	55
1.76	1.75	1.72	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.56	1.53	1.50	1.48	1.44	1.41	1.39	60
1.75	1.73	1.71	1.69	1.67	1.65	1.63	1.60	1.58	1.56	1.54	1.52	1.49	1.46	1.42	1.39	1.37	65
1.74	1.72	1.70	1.67	1.65	1.64	1.62	1.59	1.57	1.55	1.53	1.50	1.47	1.45	1.40	1.37	1.35	70
1.72	1.70	1.68	1.65	1.63	1.62	1.60	1.57	1.54	1.52	1.51	1.48	1.45	1.43	1.38	1.35	1.32	80
1.70	1.69	1.66	1.64	1.62	1.60	1.59	1.55	1.53	1.51	1.49	1.46	1.43	1.41	1.36	1.32	1.30	90
1.69	1.68	1.65	1.63	1.61	1.59	1.57	1.54	1.52	1.49	1.48	1.45	1.41	1.39	1.34	1.31	1.28	100
1.67	1.65	1.63	1.60	1.58	1.57	1.55	1.52	1.49	1.47	1.45	1.42	1.39	1.36	1.31	1.27	1.25	125
1.66	1.64	1.61	1.59	1.57	1.55	1.53	1.50	1.48	1.45	1.44	1.41	1.37	1.34	1.29	1.25	1.22	150
1.64	1.62	1.60	1.57	1.55	1.53	1.52	1.48	1.46	1.43	1.41	1.39	1.35	1.32	1.26	1.22	1.19	200
1.62	1.61	1.58	1.55	1.53	1.51	1.50	1.46	1.43	1.41	1.39	1.36	1.32	1.30	1.23	1.19	1.15	300
1.61	1.59	1.56	1.54	1.52	1.50	1.48	1.45	1.42	1.40	1.38	1.34	1.30	1.28	1.21	1.16	1.11	500
1.60	1.58	1.55	1.53	1.51	1.49	1.47	1.44	1.41	1.38	1.36	1.33	1.29	1.26	1.19	1.13	1.08	1000
1.59	1.57	1.54	1.52	1.50	1.48	1.46	1.42	1.39	1.37	1.35	1 32	1.27	1.24	1.17	1.11	1.00	∞

TABLE VII

Critical values for the F test at one-sided

Parameter f_2 of the denominator	Parameter value (degree of freedom) f_1 of the numerator																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	The values in the first row are to be multiplied by 10																	
1	405	500	540	563	576	586	593	598	602	606	608	611	613	614	616	617	618	619
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2.	27.9	27.7	27.5	27.3	27.2	27.1	27.1	27.0	26.9	26.9	26.8	26.8	26.8
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4	14.3	14.2	14.2	14.2	14.1	14.1
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.82	9.77	9.72	9.68	9.64	9.61
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31	6.27	6.24	6.21
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	6.91	5.81	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.00	4.96	4.92	4.89	4.86
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07	3.03	2.99	2.96	2.93
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98	2.94	2.91	2.88
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02	2.97	2.93	2.89	2.86	2.83
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93	2.89	2.85	2.82	2.79
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85	2.81	2.78	2.75
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.90	2.86	2.82	2.78	2.74	2.72
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.87	2.82	2.78	2.75	2.71	2.68
28	7.64	5.45	4.57	4.07	3.75	3.54	3.36	3.23	3.12	3.03	2.96	2.90	2.84	2.79	2.75	2.72	2.68	2.65
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.81	2.77	2.73	2.69	2.66	2.63
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.86	2.80	2.74	2.70	2.66	2.62	2.58	2.55
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.82	2.76	2.70	2.66	2.62	2.58	2.55	2.51
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.79	2.72	2.67	2.62	2.58	2.54	2.51	2.48
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.75	2.69	2.64	2.59	2.55	2.51	2.48	2.45
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.70	2.64	2.59	2.54	2.50	2.46	2.43	2.40
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.68	2.62	2.56	2.52	2.47	2.44	2.40	2.37
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.66	2.60	2.54	2.50	2.45	2.42	2.38	2.35
48	7.19	5.08	4.22	3.74	3.43	3.20	3.04	3.91	2.80	2.72	2.64	2.58	2.53	2.48	2.44	2.40	2.37	2.33
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.63	2.56	2.51	2.46	2.42	2.38	2.35	2.32
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.47	2.42	2.38	2.34	2.31	2.28
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35	2.31	2.28	2.25
65	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61	2.53	2.47	2.42	2.37	2.33	2.29	2.26	2.23
70	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35	2.31	2.27	2.23	2.20
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27	2.23	2.20	2.17
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29	2.24	2.21	2.17	2.14
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.26	2.22	2.19	2.15	2.12
125	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.39	2.33	2.28	2.23	2.19	2.15	2.11	2.08
150	6.81	4.75	3.92	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.37	2.31	2.25	2.20	2.16	2.12	2.09	2.06
200	6.76	4.75	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.22	2.17	2.13	2.09	2.06	2.02
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.31	2.24	2.19	2.14	2.10	2.06	2.03	1.99
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.28	2.22	2.17	2.12	2.07	2.04	2.00	1.97
1000	6.66	4.63	3.80	3.34	3.04	2.82	2,66	2.53	2.43	2.34	2.27	2.20	2.15	2.10	2.06	2.02	1.98	1.95
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.13	2.08	2.04	2.00	1.97	1.93

E.g. If the parameters are $f_1 = 8$ and $f_2 = 20$, then $P(F \leqslant 3.56) = 99\%$.

(cont.)

level 99%, or two-sided level 98%

Parameter value (degree of freedom) f_1 of the numerator																	Parameter f_2 of the denominator
19	20	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞	
The values in the first row are to be multiplied by 10																	
620	621	622	623	624	625	626	628	629	630	630	631	633	633	635	636	637	1
99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	2
26.7	26.7	26.6	26.6	26.6	26.5	26.5	26.5	26.4	26.4	26.4	26.3	26.3	26.2	26.2	26.1	26.1	3
14.0	14.0	14.0	13.9	13.9	13.9	13.8	13.8	13.7	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5	4
9.58	9.55	9.51	9.47	9.43	9.40	9.38	9.33	9.29	9.26	9.24	9.20	9.16	9.13	9.08	9.04	9.02	5
7.42	7.40	7.35	7.31	7.28	7.25	7.23	7.18	7.14	7.11	7.09	7.06	7.01	6.99	6.93	6.90	6.88	6
6.18	6.16	6.11	6.07	6.04	6.02	5.99	5.94	5.91	5.88	5.86	5.82	5.78	5.75	5.70	5.67	5.65	7
5.38	5.36	5.32	5.28	5.25	5.22	5.20	5.15	5.12	5.09	5.07	5.03	4.99	4.96	4.91	4.88	4.86	8
4.83	4.81	4.77	4.73	4.70	4.67	4.65	4.60	4.57	4.54	4.52	4.48	4.44	4.42	4.36	4.33	4.31	9
4.43	4.41	4.36	4.33	4.30	4.27	4.25	4.20	4.17	4.14	4.12	4.08	4.04	4.01	3.96	3.93	3.91	10
4.12	4.10	4.06	4.02	3.99	3.96	3.94	3.89	3.86	3.83	3.81	3.78	3.73	3.71	3.66	3.62	3.60	11
3.88	3.86	3.82	3.78	3.75	3.72	3.70	3.65	3.62	3.59	3.57	3.54	3.49	3.47	3.41	3.38	3.36	12
3.69	3.66	3.62	3.59	3.56	3.53	3.51	3.46	3.43	3.40	3.38	3.34	3.30	3.27	3.22	3.19	3.17	13
3.53	3.51	3.46	3.43	3.40	3.37	3.35	3.30	3.27	3.24	3.22	3.18	3.14	3.11	3.06	3.03	3.00	14
3.40	3.37	3.33	3.29	3.26	3.24	3.21	3.17	3.13	3.10	3.08	3.05	3.00	2.98	2.92	2.89	2.87	15
3.28	3.26	3.22	3.18	3.15	3.12	3.10	3.05	3.02	2.99	2.97	2.93	2.89	2.86	2.81	2.78	2.75	16
3.18	3.16	3.12	3.08	3.05	3.03	3.00	2.96	2.92	2.89	2.87	2.83	2.79	2.76	2.71	2.68	2.65	17
3.10	3.08	3.03	3.00	2.97	2.94	2.92	2.87	2.84	2.81	2.78	2.75	2.70	2.68	2.62	2.59	2.57	18
3.03	3.00	2.96	2.92	2.89	2.87	2.84	2.80	2.76	2.73	2.71	2.67	2.63	2.60	2.55	2.51	2.49	19
2.96	2.94	2.90	2.86	2.83	2.80	2.78	2.73	2.69	2.67	2.64	2.61	2.56	2.54	2.48	2.44	2.42	20
2.90	2.88	2.84	2.80	2.77	2.74	2.72	2.67	2.64	2.61	2.58	2.55	2.50	2.48	2.42	2.38	2.36	21
2.85	2.83	2.78	2.75	2.72	2.69	2.67	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.36	2.33	2.31	22
2.80	2.78	2.74	2.70	2.67	2.64	2.62	2.57	2.54	2.51	2.48	2.45	2.40	2.37	2.32	2.28	2.26	23
2.76	2.74	2.70	2.66	2.63	2.60	2.58	2.53	2.49	2.46	2.44	2.40	2.36	2.33	2.27	2.24	2.21	24
2.72	2.70	2.66	2.62	2.59	2.56	2.54	2.49	2.45	2.42	2.40	2.36	2.32	2.29	2.23	2.19	2.17	25
2.69	2.66	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.39	2.36	2.33	2.28	2.25	2.19	2.16	2.13	26
2.66	2.63	2.59	2.55	2.52	2.49	2.47	2.42	2.38	2.35	2.33	2.29	2.25	2.22	2.16	2.12	2.10	27
2.63	2.60	2.56	2.52	2.49	2.46	2.44	2.39	2.35	2.32	2.30	2.26	2.22	2.19	2.13	2.09	2.06	28
2.60	2.57	2.53	2.49	2.46	2.44	2.41	2.36	2.33	2.30	2.27	2.23	2.19	2.16	2.10	2.06	2.03	29
2.57	2.55	2.51	2.47	2.44	2.41	2.39	2.34	2.30	2.27	2.25	2.21	2.16	2.13	2.07	2.03	2.01	30
2.53	2.50	2.46	2.42	2.39	2.36	2.34	2.29	2.25	2.22	2.20	2.16	2.11	2.08	2.02	1.98	1.96	32
2.49	2.46	2.42	2.38	2.35	2.32	2.30	2.25	2.21	2.18	2.16	2.12	2.07	2.04	1.98	1.94	1.91	34
2.45	2.43	2.38	2.35	2.32	2.29	2.26	2.21	2.17	2.14	2.12	2.08	2.03	2.00	1.94	1.90	1.87	36
2.42	2.40	2.35	2.32	2.28	2.26	2.23	2.18	2.14	2.11	2.09	2.05	2.00	1.97	1.90	1.86	1.84	38
2.39	2.37	2.33	2.29	2.26	2.23	2.20	2.15	2.11	2.08	2.06	2.02	1.97	1.94	1.87	1.83	1.80	40
2.37	2.34	2.30	2.26	2.23	2.20	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.91	1.85	1.80	1.78	42
2.35	2.32	2.28	2.24	2.21	2.18	2.15	2.10	2.06	2.03	2.01	1.97	1.92	1.89	1.82	1.78	1.75	44
2.33	2.30	2.26	2.22	2.19	2.16	2.13	2.08	2.04	2.01	1.99	1.95	1.90	1.86	1.80	1.75	1.73	46
2.31	2.28	2.24	2.20	2.17	2.14	2.12	2.06	2.02	1.99	1.97	1.93	1.88	1.84	1.78	1.73	1.70	48
2.29	2.27	2.22	2.18	2.15	2.12	2.10	2.05	2.01	1.97	1.95	1.91	1.86	1.82	1.76	1.71	1.68	50
2.25	2.23	2.18	2.15	2.11	2.08	2.06	2.01	1.97	1.93	1.91	1.87	1.81	1.78	1.71	1.67	1.64	55
2.22	2.20	2.15	2.12	2.08	2.05	2.03	1.90	1.94	1.90	1.88	1.84	1.78	1.75	1.68	1.63	1.60	60
2.20	2.17	2.13	2.09	2.06	2.03	2.00	1.95	1.91	1.88	1.85	1.81	1.75	1.72	1.65	1.60	1.57	65
2.18	2.15	2.11	2.07	2.03	2.01	1.98	1.93	1.89	1.85	1.83	1.78	1.73	1.70	1.62	1.57	1.54	70
2.14	2.12	2.07	2.03	2.00	1.97	1.94	1.89	1.85	1.81	1.79	1.75	1.69	1.66	1.58	1.53	1.49	80
2.11	2.09	2.04	2.00	1.97	1.94	1.92	1.86	1.82	1.79	1.76	1.72	1.66	1.62	1.54	1.49	1.46	90
2.09	2.07	2.02	1.98	1.94	1.92	1.89	1.84	1.80	1.76	1.73	1.69	1.63	1.60	1.52	1.47	1.43	100
2.05	2.03	1.98	1.94	1.91	1.88	1.85	1.80	1.76	1.72	1.69	1.65	1.59	1.55	1.47	1.41	1.37	125
2.03	2.00	1.96	1.92	1.88	1.85	1.83	1.77	1.73	1.69	1.66	1.62	1.56	1.52	1.43	1.38	1.33	150
2.00	1.97	1.93	1.89	1.85	1.82	1.79	1.74	1.69	1.66	1.63	1.58	1.52	1.48	1.39	1.33	1.28	200
1.97	1.94	1.89	1.85	1.82	1.79	1.76	1.71	1.66	1.62	1.59	1.55	1.48	1.44	1.35	1.28	1.22	300
1.94	1.92	1.87	1.83	1.79	1.76	1.74	1.68	1.63	1.60	1.56	1.52	1.45	1.41	1.31	1.23	1.16	500
1.92	1.90	1.85	1.81	1.77	1.74	1.72	1.66	1.61	1.57	1.54	1.50	1.43	1.38	1.28	1.19	1.11	1000
1.90	1.88	1.83	1.79	1.76	1.72	1.70	1.64	1.59	1.55	1.52	1.47	1.40	1.36	1.25	1.15	1.00	∞

TABLE VII (cont.)

Critical values of the F test for the one-sided level 99.9%, or two-sided level 99.8%

Parameter f_2 of the denominator	Parameter (degree of freedom) f_1 of the numerator																	
	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
	The values in the first row are to be multiplied by 1000																	
1	405	500	540	562	576	586	593	598	602	606	616	621	626	630	633	635	636	637
2	998	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3	168	148	141	137	135	133	132	131	130	129	127	126	125	125	124	124	124	124
4	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48 .5	48.0	46.8	46.1	45.4	44.9	44.5	44.3	44.1	44.0
5	47.0	36.6	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	25.9	25.4	24.9	24.4	24.1	23.9	23.8	23.8
6	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7	18.4	17.6	17.1	16.7	16.3	16.0	15.9	15.8	15.8
7	29.2	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.3	14.1	13.3	12.9	12.5	12.2	11.9	11.8	11.7	11.7
8	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8	11.5	10.8	10.5	10.1	9.80	9.57	9.46	9.39	9.34
9	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1	9.89	9.24	8.90	8.55	8.26	8.04	7.93	7.86	7.81
10	21.0	14.9	12.6	11.3	10.5	9.92	9.52	9.20	8.96	8.75	8.13	7.80	7.47	7.19	6.98	6.87	6.81	6.76
11	19.7	13.8	11.6	10.4	9.58	9.05	8.66	8.35	8.12	7.92	7.32	7.01	6.68	6.41	6.21	6.10	6.04	6.00
12	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48	7.29	6.71	6.40	6.09	5.83	5.63	5.52	5.46	5.42
13	17.8	12.3	10.2	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.23	5.93	5.62	5.37	5.17	5.07	5.01	4.97
14	17.1	11.8	9.73	8.62	7.92	7.43	7.08	6.80	6.58	6.40	5.85	5.56	5.25	5.00	4.80	4.70	4.64	4.60
15	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.53	5.25	4.95	4.70	4.51	4.41	4.35	4.31
16	16.1	11.0	9.00	7.94	7.27	6.81	6.46	6.19	5.98	5.81	5.27	4.99	4.70	4.45	4.26	4.16	4.10	4.06
17	15.7	10.7	8.73	7.68	7.02	6.56	6.22	5.96	5.75	4.58	5.05	4.78	4.48	4.24	4.05	3.95	3.89	3.85
18	15.4	10.4	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	4.87	4.59	4.30	4.06	4.87	3.77	3.71	3.67
19	15.1	10.2	8.28	7.26	6.61	6.18	5.84	5.59	5.39	5.22	4.70	4.43	4.14	3.90	3.71	3.61	3.55	3.51
20	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.56	4.29	4.01	3.77	3.58	3.48	3.42	3.38
22	14.4	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.32	4.06	3.77	3.53	3.34	3.25	3.19	3.15
24	14.0	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.14	3.87	3.59	3.35	3.16	3.07	3.01	2.97
26	13.7	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	3.99	3.72	3.45	3.20	3.01	2.92	2.86	2.82
28	13.5	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	3.86	3.60	3.32	3.08	2.89	2.79	2.73	2.70
30	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	3.75	3.49	3.22	2.98	2.79	2.69	2.63	2.59
40	12.6	8.25	6.60	5.70	5.13	4.73	4.43	4.21	4.02	3.87	3.40	3.15	2.87	2.64	2.44	2.34	2.28	2.23
50	12.2	7.95	6.34	5.46	4.90	4.51	4.22	4.00	3.82	3.67	3.20	2.95	2.68	2.44	2.24	2.14	2.07	2.03
60	12.0	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69	3.54	3.08	2.83	2.56	2.31	2.11	2.01	1.93	1.89
80	11.7	7.54	5.97	5.13	4.58	4.21	3.92	3.70	3.53	3.39	2.93	2.68	2.40	2.16	1.95	1.84	1.77	1.72
100	11.5	7.41	5.85	5.01	4.48	4.11	3.83	3.61	3.44	3.30	2.84	2.59	2.32	2.07	1.87	1.75	1.68	1.62
200	11.2	7.15	5.64	4.81	4.29	3.92	3.65	3.43	3.26	3.12	2.67	2.42	2.15	1.90	1.68	1.55	1.46	1.39
500	11.0	7.01	5.51	4.69	4.18	3.82	3.54	3.33	3.16	3.02	2.58	2.33	2.05	1.80	1.57	1.43	1.32	1.23
∞	10.8	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.51	2.27	1.99	1.73	1.49	1.34	1.12	1.00

TABLE VIII

CRITICAL VALUES OF THE t (STUDENT)TEST AT ONE-(TWO-)SIDED LEVELS

One-sided Two-sided Degrees of freedom	97.5% 95.0%	99.5% 99.0%	99.95% 99.9%
1	12.71	63.66	636.6
2	4.303	9.925	31.60
3	3.182	5.841	12.94
4	2.776	4.604	8.610
5	2.571	4.032	6.859
6	2.447	3.707	5.959
7	2.365	3.499	5.405
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.106	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073
16	2.120	2.921	4.015
17	2.110	2.898	3.965
18	2.101	2.878	3.922
19	2.093	2.861	3.883
20	2.086	2.845	3.850
21	2.080	2.831	3.819
22	2.074	2.819	3.792
23	2.069	2.807	3.767
24	2.064	2.797	3.745
25	2.060	2.787	3.725
26	2.056	2.779	3.707
27	2.052	2.771	3.690
28	2.048	2.763	3.674
29	2.045	2.756	3.659
30	2.042	2.750	3.646
40	2.021	2.704	3.551
50	2.009	2.678	3.495
60	2.000	2.660	3.460
80	1.990	2.639	3.415
100	1.984	2.626	3.389
200	1.972	2.601	3.339
500	1.965	2.586	3.310
∞	1.960	2.576	3.291

$$P\left(\overline{X} - t_0 \frac{s_n}{\sqrt{n}} < \mu < \overline{X} + t_0 \frac{s_n}{\sqrt{n}}\right) = P(|t| < t_0)$$

where μ is the expected value of the Gaussian variable, $\overline{X}$ is the sample mean, s^2 the empirical variance, t the Student variable, t_0 the value found in the Table; e.g., if $f = 9$, $P(|t| < 2.262) = 95\%$.

TABLE IX

QUANTILES OF THE χ^2 VARIABLE, I.E., CRITICAL VALUES TO THE χ^2 TEST

Deegres of freedom f	Probability expressed in percent						
	90.0	95.0	97.5	99.0	99.5	99.9	99.95
1	2.71	3.84	5.02	6.63	7.88	10.8	12.1
2	4.61	5.99	7.38	9.21	10.6	13.8	15.2
3	6.25	7.81	9.35	11.3	12.8	16.3	17.7
4	7.78	9.49	11.1	13.3	14.9	18.5	20.0
5	9.24	11.1	12.8	15.1	16.7	20.5	22.1
6	10.6	12.6	14.4	16.8	18.5	22.5	24.1
7	12.0	14.1	16.0	18.5	20.3	24.3	26.0
8	13.4	15.5	17.5	20.1	22.0	26.1	27.9
9	14.7	16.9	19.0	21.7	23.6	27.9	29.7
10	16.0	18.3	20.5	23.2	25.2	29.6	31.4
11	17.3	19.7	21.9	24.7	26.8	31.3	33.1
12	18.5	21.0	23.3	26.2	28.3	32.9	34.8
13	19.8	22.4	24.7	27.7	29.8	34.5	36.5
14	21.1	23.7	26.1	29.1	31.3	36.1	38.1
15	22.3	25.0	27.5	30.6	32.8	37.7	39.7
16	23.5	26.3	28.8	32.0	34.3	39.3	41.3
17	24.8	27.6	30.2	33.4	35.7	40.8	42.9
18	26.0	28.9	31.5	34.8	37.2	42.3	44.4
19	27.2	30.1	32.9	36.2	38.6	43.8	46.0
20	28.4	31.4	34.2	37.6	40.0	45.3	47.5
21	29.6	32.7	35.5	38.9	41.4	46.8	49.0
22	30.8	33.9	36.8	40.3	42.8	48.3	50.5
23	32.0	35.2	38.1	41.6	44.2	49.7	52.0
24	33.2	36.4	39.4	43.0	45.6	51.2	53.5
25	34.4	37.7	40.6	44.3	46.9	52.6	54.9
26	35.6	38.9	41.9	45.6	48.3	54.1	56.4
27	36.7	40.1	43.2	47.0	49.6	55.5	57.9
28	37.9	41.3	44.5	48.3	51.0	56.9	59.3
29	39.1	42.6	45.7	49.6	52.3	58.3	60.7
30	40.3	43.8	47.0	50.9	53.7	59.7	62.2

$P(\chi^2 \leqslant \chi_0^2)$, where χ^2 is the observed (calculated) value, χ_0^2 the value found in the Table. E.g., if $f = 11$, $P(\chi^2 \leqslant 19.4) = 95\%$.

TABLE X

TABLE TO KOLMOGOROV'S ONE SAMPLE TEST
Two-sided alternatives at levels 95% and 99%

n	0.95	0.99	n	0.95	0.99
8	0.4543	0.5419	21	0.2827	0.3443
9	0.4300	0.5133	22	0.2809	0.3367
10	0.4093	0.4889	23	0.2749	0.3295
			24	0.2693	0.3229
11	0.3912	0.4677	25	0.2640	0.3166
12	0.3754	0.4491			
13	0.3614	0.4325	26	0.2591	0.3106
14	0.3489	0.4176	27	0.2544	0.3050
15	0.3376	0.4042	28	0.2499	0.2997
			29	0.2457	0.2947
16	0.3273	0.3920	30	0.2417	0.2899
17	0.3180	0.3809			
18	0.3094	0.3706	35	0.2243	0.2690
19	0.3014	0.361	40	0.2101	0.2521
20	0.2941	0.3524	45	0.1984	0.2380
			50	0.1884	0.2260

If $D_n = \max | F_n(x) - F(x)|$ is equal to or larger than the critical value found in the Table, then the hypothesis that the true distribution function is $F(x)$, will be rejected.

TABLE XI

CRITICAL VALUES TO THE KOLMOGOROV–SMIRNOV TWO-SAMPLE TEST. THE CASE OF EQUAL SAMPLE SIZES $m = n$, AGAINST TWO-SIDED ALTERNATIVES AT LEVEL 95% AND 99%

n	0.95	0.99	n	0.95	0.99	n	0.95	0.99
7	6	6	15	8	9	23	10	11
8	6	7	16	8	10	24	10	12
9	6	7	17	8	10	25	10	12
10	7	8	18	9	10	26	10	12
11	7	8	19	9	10	27	10	12
12	7	8	20	9	11	28	11	13
13	7	9	21	9	11	29	11	13
14	8	9	22	9	11	30	11	13

If $nD_{n,n} = n \max\limits_{(x)} | F_n(x) - G_n(x)|$ is equal to or larger than the value found in the Table, then the hypothesis of the agreement of the two distribution functions will be rejected.

TABLE XII

CRITICAL VALUES OF THE WILCOXON TWO-SAMPLE TEST AT ONE-SIDED (TWO-SIDED) LEVEL 97.5% (95%)

m n	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	77
14	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83
15	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105
18	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

CRITICAL VALUES OF THE WILCOXON TWO-SAMPLE TEST AT ONE-SIDED (TWO-SIDED) LEVEL 99.5% (99.0%)

m n	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	7	10	13	17	20	24	27	31	34	38	42	45	49	53	57	60
14	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	10	15	19	24	29	34	39	44	49	55	60	65	70	75	81	86
18	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	12	17	22	28	33	39	45	51	57	63	69	74	81	87	93	99
20	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

Hint. If the observed value of the U statistic given in 4.3.7.5 is less than $\frac{nm}{2}$, then U itself, otherwise the value $(nm - U)$ is to be compared with the critical value found in the Table. If the actual value of the statistic does not exceed the critical value, then the hypothesis of the agreement of the two distribution functions is to be rejected at the following level: The first part of the Table in the case of one-sided alternative to the level $1 - \alpha = 0.975 = 97.5\%$, while in the case of two-sided alternative to the level $1 - \alpha = 0.95 = 95\%$ corresponds. The critical values of the second part of the Table belong to the level $1 - \alpha = 0.995 = 99.5\%$ (one-sided) and to the level $1 - \alpha = 0.99 = 99\%$ (two-sided), respectively.

BIBLIOGRAPHY

TEXTBOOKS

[A 1] Anderson, T. W.: An Introduction to Multivariate Statistical Analysis. Wiley, New York, 1958.

[A 2] Barlow, R. E.—Proschan, F.: Mathematical Theory of Reliability. Wiley, New York, 1965.

[A 3] Bowker, A. H.—Lieberman, G. J.: Engineering Statistics. Prentice Hall, Englewood Cliffs, 1959.

[A 4] Buckland, W. R.: Statistical Assessment of the Life Characteristics. Griffin, London, 1964.

[A 5] Burr, I. W.: Engineering Statistics and Quality Control. McGraw-Hill, New York, 1953.

[A 6] Cavé, R.: Le contrôle statistique des fabrications. 3^{e}. éd. Eyrolles, Paris 1966.

[A 7] Cowden, D. J.: Statistical Methods in Quality Control. Prentice-Hall, Englewood Cliffs, 1957.

[A 8] Cox, D. R.: Renewal Theory, Methuen, London, 1962.

[A 9] Deming, W. E.: Some Theory of Sampling. Wiley, New York, 1950.

[A 10] Deming, W. E.: Statistical Adjustment of Data, Dover Publ., New York 1964.

[A 11] Dobben de Bruyn, C. S. van: Cumulative Sum Tests: Theory and Practice, Griffin, London, 1968.

[A 12] Duncan, A. J.: Quality Control and Industrial Statistics. Irwin, Homewood 1955.

[A 13] Ehrenfeld, S.—Littauer, S. B.: Introduction to Statistical Methods. McGraw-Hill, New York, 1964.

[A 14] Eisenhart, C.—Hastay, M. W.—Wallis, W. A.: Techniques of Statistical Analysis. McGraw-Hill, New York, 1947.

[A 15] Ezekiel, M.—Fox, K. A.: Methods of Correlation and Regression Analysis, Wiley, New York, 1959.

[A 16] Ferguson, T.: Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York, 1967.

[A 17] Fraser, D. A. S.: Nonparametric Methods in Statistics. Wiley, New York, 1957.

[A 18] Freeman, H. A.—Friedman, J.—Mosteller, F.—Wallis, W. A.: Sampling Inspection. McGraw-Hill, 1948.

[A 19] Gnedenko, B. V.—Belyayev, Yu. K.—Solovyev, A. D.: Mathematical Methods of Reliability Theory. Academic Press, New York, 1969.

[A 20] Grenander, U.—Rosenblatt, M.: Statistical Analysis of Stationary Time Series, Wiley, New York, 1957.

[A 21] Grant, E. L.: Statistical Quality Control. 2nd ed., McGraw-Hill, New York, 1952.

[A 22] Haight, F.: Handbook of the Poisson Distribution. Wiley, New York, 1967.

[A 23] Hájek, J.—Šidák, Z.: Theory of Rank Tests. Academia, Prague, 1967.

[A 24] Hald, A.: Statistical Theory with Engineering Applications. Wiley, New York, 1952.

[A 25] Ireson, W. G. (ed.): Reliability Handbook. McGraw-Hill, New York, 1952.

[A 26] Johnson, W. L.—Leone, F. C.: Statistics and Experimental Design I. Wiley, New York, 1964.

[A 27] Juran, J. M. (ed.): Quality Control Handbook. 2nd ed., McGraw-Hill, New York, 1962.

[A 28] Kendall, M. G.—Stuart, A.: The Advanced Theory of Statistics. I, II, III, Griffin, London, 1958, 1961, 1966.

[A 29] Kolmogorov, A. N.: Foundations of the Theory of Probability. Chelsea Publishing Company, New York, 2nd ed. 1956.

[A 30] Lehmann, E. L.: Testing Statistical Hypotheses. Wiley, New York, 1959.

[A 31] Miller, I.—Freund, J. E.: Probability and Statistics for Engineers. Prentice Hall, Englewood Cliffs, 1965.

[A 32] Parzen, E.: Stochastic processes. Holden-Day, San Francisco, 1963.

[A 33] Pearson, E. S.: Applications of Statistical Methods (BS-600: 1935).

[A 34] Plackett, R. L.: Principles of Regression Analysis. Clarendon Press, Oxford, 1960.

[A 35] Rickmers, A. D.—Todd, H. N.: Statistics. An Introduction. McGraw-Hill, New York, 1967.

[A 36] Sarhan, A. E.—Greenberg, B. G.: Contributions to Order Statistics. Wiley, New York, 1962.

[A 37] Scheffé, H.: The Analysis of Variance. Wiley, New York, 1959.

[A 38] Schindowski, E.—Schürz, O.: Statistische Qualitätskontrolle. Verlag Technik, Berlin, 1965.

[A 39] Schmetterer, L.: Einführung in die mathematische Statistik. 2nd ed. Springer, Wien—New York, 1966.

[A 40] Shewhart, W.: Economic Control of Quality of Manufactured Product. Van Nostrand, Princeton, 1931.

[A 41] Stange, K.: Angewandte Statistik I—II. Springer, Berlin, 1970—71.

[A 42] Storm, R.: Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle. 3. Aufl. VEB Fachbuchverlag, Leipzig, 1969.

[A 43] Störmer, H.: Mathematische Theorie der Zuverlässigkeit. Oldenburg, München, 1970.

[A 44] Takács, L.: Stochastic Processes. Problems and Solutions. Methuen, London, 1960.

[A 45] Tucker, H.: A Graduate Course in Probability. Academic Press, New York, 1967.

[A 45a] Uhlmann, W.: Kostenoptimale Prüfpläne: Tabellen, Praxis und Theorie eines Verfahrens der statistischen Qualitätskontrolle, Physica, Würzburg, 1969.

[A 45b] UHLMANN, W. Statistische Qualitätskontrolle. Teubner, Stuttgart, 1966.

[A 46] VINCZE, I.: Mathematische Statistik mit industriellen Anwendungen, Akadémiai Kiadó, Budapest, 1971.

[A 47] VINCZE, I. (ed.): Statisztikai minőségellenőrzés: Az ipari minőségellenőrzés matematikai statisztikai módszerei (Statistical Quality Control: The Mathematical Statistical Methods of Industrial Quality Control), Közgazdasági és Jogi Könyvkiadó, Budapest, 1958.

[A 48] WALD, A.: Sequential Analysis. Wiley, New York, 1947.

[A 49] WALD, A.: Statistical Decision Functions, Wiley, New York, 1961.

[A 50] ZELEN, M. (ed.): Statistical Theory of Reliability, University of Wisconsin Press, Madison, 1963.

ARTICLES

[B 1] BARLOW, R. E.: Some recent developments in reliability theory. Operations Research Center, ORC 68—19 (1968), University of California, Berkeley.

[B 2] BARLOW, R. E.—MARSHALL, A. W.: Bounds for distributions with monotone, hazard rate. Ann. Math. Statist. I—II 35 (1964) 1234—1274.

[B 3] BARLOW, R. E.—MARSHALL, A. W.—PROSCHAN, F.: Properties of probability distributions with monotone, hazard rate. Ann. Math. Statist. 34 (1963) 375—389.

[B 4] BARLOW, R. E.—PROSCHAN, F.: A note on tests for monotone failure rate based on incomplete data. Ann. Math. Statist. 40 (1969) 595—600.

[B 5] BARNARD, G. A.: Sampling inspection and statistical decisions. Journ. Roy. Statist. Soc. B 16 (1954) 151—174.

[B 6] BARNARD, G. A.: Control charts and stochastic processes. Journ. Roy. Staist. Soc. B 21 (1959) 239—271.

[B 7] BATHER, J. A.: Control charts and the minimization of costs. Journ. Roy. Statist. Soc. B 25 (1963) 49—80.

[B 8] BIRNBAUM, Z. W.: On the power of a one-sided test of fit for continuous probability functions. Ann. Math. Statist. 24 (1953) 484—489.

[B 9] BIRNBAUM, Z. W.—ESARY, J. D.—MARSHALL, A. W.: A stochastic characterisation of wear-out for components and systems. Ann. Math. Statist. 37 (1966) 816—855.

[B 10] BIRNBAUM, Z. W.—ESARY, J. D.—SAUNDERS, S. C.: Multicomponent systems and structures and their reliability. Technometrics 3 (1961) 55—57.

[B 11] BIRNBAUM, Z. W.—TINGEY, F. H.: One-sided confidence contours for probability distribution functions. Ann. Math. Statist. 22 (1951) 592—596.

[B 12] BREAKWELL, J. V.: Economically optimum sampling tests. Journ. Amer. Statist. Assoc. 51 (1956) 243—256.

[B 13] BURR, I. W.: New methods for approving machine and process setting. Part I. Industrial Quality Control 5/4 (1949) 12—18; Part II. *ib.* 6/3 (1950) 13—16.

[B 14] CAVÉ, R.: Contrôle statistique par calibres modifiés d'efficacité optima. Comptes rendus Acad. Sci. Paris 235 (1952) 935—937.

[B 15] CHAMPERNOWNE, D. G.: The economics of sequential sampling procedures for defectives. Applied Statist. 2 (1953) 118—130.

[B 16] CHAPMAN, D. S.: A comparative study of several one-sided goodness of fit tests. Ann. Math. Statist. 29 (1958) 655—674.

[B 17] CRAIG, C. O.: Control charts versus the analysis of variance in process control by variables. Industrial Quality Control 3/4 (1947) 14—18.

[B 18] DEMING, W. E.: What happened in Japan? Industrial Quality Control 24/2 (1967) 89—93.

[B 19] DIANANDA, P. H.: Some probability limit theorems with statistical applications. Proc. Cambr. Phil. Soc. 49 (1953) 239—246.

[B 20] DOKSUM, K.: Asymptotically optimal statistics in some models with increasing failure rate average. Ann. Math. Statist. 38 (1967) 1731—1739.

[B 21] ÉLTETŐ, Ö.—SARKADI, K.: On preliminary sampling with special reference to the microbiological control of tomato concentrates. Biometrics. 16 (1960) 339—347.

[B 22] EPSTEIN, B.: Estimation from life test data. Technometrics 2 (1960) 448—454.

[B 23] EPSTEIN, B.: Statistical life test acceptance procedures. Technometrics 2 (1960) 435—446.

[B 24] EPSTEIN, B.—SOBEL, M.: Sequential life tests in the exponential case. Ann. Math. Statist. 26 (1957) 82—93.

[B 25] ESARY, J. D.—PROSCHAN, F.: Coherent structures on non-identical components. Technometrics 5 (1963) 191—209.

[B 26] ESARY, J. D.—PROSCHAN, F.—WALKUP, D. W.: A multivariate notion of association with a reliability application. Boeing Stat. Research Laboratories DI-82-0567 (1966).

[B 27] FONTÁNYI, Á.—SARKADI, K.—VAS É.: Application of the theory of order statistics in the statistical quality control. Publ. Math. Inst. Hung. Acad. Sci. 2 (1953) 307—324 (in Hungarian).

[B 28] FREUND, R. A.: Acceptance control charts. Industrial Quality Control, 14/4 (1957) 13—23.

[B 29] FREUND, R. A.: Variables control chart. Industrial Quality Control 16/11 (1960) 35—41.

[B 30] GNEDENKO, B. V.: On statistical methods in reliability theory. European Meeting on Statistics, Econometrics and Management Science, Amsterdam, 1968.

[B 31] GRENANDER, U.: On the theory of mortality measurement. I—II. Skand. Akt.-tidskr. 39 (1956) 125—153.

[B 32] GRIFFITHS, B.—RAO, A. G.: An application of least cost acceptance samplings schemes. Unternehmensforschung (Operations Research) 9 (1964) 8—17.

[B 33] GUTHRIE, D. JR.—M. V. JOHNS, JR.: Bayes acceptance sampling procedures for large lots. Ann. Math. Statist. 30 (1959) 896—925.

[B 34] HALD, A.: Bayesian single sampling attribute plans for discrete distribution. Math. Phys. Skr. Dan. Vid. Selsk. 3 (1965).

[B 35] HALD, A.: The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs. Technometrics 2 (1960) 275—340.

[B 36] Hald, A.: Asymptotic properties of Bayesian single sampling plans. Journ. Roy. Statist. Soc. B 29 (1967) 162—173.

[B 37] Hald, A.: Bayesian single sampling attribute plans for continuous prior distribution. Technometrics 10 (1968) 667—683.

[B 38] Hald, A.—Thyregod, P.: The composite operating characteristic under normal and tightened sampling inspection by attributes. Bull. Intern. Statist. Inst. 41 (1966) 517—539.

[B 39] Hamaker, H. C.—Strik, R. van: The efficiency of double sampling for attributes. Journ. Amer. Statist. Assoc. 50 (1955) 830—849.

[B 40] Hill, I. D.: Sampling inspection and defense specification BEF-131. Journ. Roy. Statist. Soc. A 125 (1962) 31—73.

[B 41] Horshell, G.: Economical acceptance sampling schemes. Journ. Roy. Statist. Soc. A 120 (1957) 148—175.

[B 42] Howell, J. M.: Control chart for largest and smallest values. Ann. Math. Statist. 20 (1949) 305—309.

[B 43] Hunter, L. C.: Optimum checking procedures. Statistical Theory of Reliability. Ed. M. Zelen, The University of Wisconsin Press, Madison, 1963. [A 50] 95—167.

[B 44] James, A. N.: The statistician and sample testing. Bull. Inst. Physics (1958) 182—188.

[B 45] Kolmogorov, A. N.: Unbiased estimates. Amer. Math. Soc. Transl. 98 (1953).

[B 46] Liebermann, G. J.: Statistical process control and the impact of automatic process control. Technometrics 7 (1965) 283—292.

[B 47] Liebermann, G. J.—Resnikoff, G. J.: Sampling plans for inspection by variables. Journ. Amer. Statist. Assoc. 50 (1955) 457—516.

[B 47a] Lilliefors, H. W.: On the Kolgomorov–Smirnov test for normality with mean and variance unknown. Journ. Amer. Statist. Assoc. 62 (1967) 393—402.

[B 48] Mann, N.: Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions. Technometrics 10 (1968) 231—256.

[B 49] Marshall, A. W.—Proschan, F.: Maximum likelihood estimation for distributions with monotone failure rate. Ann. Math. Statist. 36 (1965) 69—77.

[B 50] Masuyama, M.: Tables of two-sided 5% and 1% control limits for individual observations of the r'th order. Sankhya 15 (1955) 291—294.

[B 51] Moore, E. F.—Shannon, C. E.: Reliable circuits using less reliable relays. Journ. Franklin Inst. 262 (1956) 191—208, 281—297.

[B 52] Neumann, J. von: Probabilistic logics and synthesis of reliable organum from unreliable components. Automata Studies 43—48, Ann. of Math. Studies No. 34, Princeton University Press, 1956.

[B 53] Oderfeld, J.: On the dual aspect of sampling plans. Coll. Math. 2 (1951) 89—97.

[B 54] Owen, D. B.: Variables sampling plans based on the normal distribution. Technometrics 9 (1967) 417—423.

[B 55] Pabst, W. R. Jr : MIL-STD-105D. Industrial quality control 20/5 (1963) 4—9.

[B 56] Pabst, W. R. Jr.: International aspects of p. c. activity. Rao, C. R. (ed.): Contributions to statistics (presented to C. Mahalanobis), Pergamon Press, Oxford, 1964.

[B 57] Page, E. S.: Continuous inspection schemes. Biometrika 41 (1954) 100 – 114 100 – 114.

[B 58] Peach, P. — Littauer, S. B.: A note on sampling inspection. Ann. Math. Statist. 17 (1946) 81 – 87.

[B 59] Pfanzagl, J.: Sampling procedures based on prior distributions and costs. Technometrics 5 (1963) 47 – 61.

[B 60] Phatak, A. G. and Bhatt, N. M.: Estimation of fraction defective in curtailed sampling plans by attributes. Technometrics 9 (1967) 219 – 228.

[B 61] Proschan, F. — Savage, I. R.: Starting a control chart. The effect of number and size of samples on the level of significance at the start of a control chart for sample mean. Industrial Quality Control. 7 (1960) 12 – 13.

[B 62] Pugh, E. L.: The best estimate of reliability in the exponential case. Oper. Res. 11/17 (1963) 57 – 61.

[B 63] Roberts, S. W.: A comparison of some control chart procedures. Technometrics 8 (1966) 411 – 430.

[B 64] Rubinstein, D.: Estimation of failure rates of systems in development. Ann. Math. Stat. 32 (1961) 924.

[B 65] Sarkadi, K.: On a rule of dualism in mathematical statistics. Acta Math. Acad. Sci. Hung. 11 (1960) 83 – 92.

[B 66] Sarkadi, K.: Estimation after selection. Studia Sci. Math. Hung. 2 (1967) 341 – 350.

[B 67] Sarkadi, K.: On testing for normality. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1966, 373 – 387.

[B 68] Sarkadi, K.: Some remarks on the estimation of fraction defective and reliability in the case of normal distribution. Studies in Mathematical Statistics. Publishing House of the Hungarian Academy of Sciences, Budapest, 1968. 133 – 137.

[B 69] Sathe, Y. S. — Varde, S. D.: Minimum variance unbiased estimation of reliability for the truncated exponential distribution. Technometrics 11 (1969) 609 – 612.

[B 70] Saunders, S.: On the determination of a safe life for classes of distributions classified by failure rate. Technometrics 10 (1968) 361 – 377.

[B 70a] Shapiro, S. S. — Francia, R. S.: Approximate analysis of variance test for normality. Journ. Amer. Statist. Assoc. 67. (1972) 215 – 216.

[B 70b] Shapiro, S. S. — Wilk, N.B.: An analysis of variance test for normality (complete samples). Biometrika 52 (1965) 591 – 611.

[B 70c] Shapiro, S. S. — Wilk, M. B. — Chen, H. J.: A comparative study of various tests for normality. Journ. Amer. Statist. Assoc. 63 (1968) 1343 – 1372.

[B 71] Shewhart, W. A.: Quality control charts. Bell System Technical Journal 5 (1926) 593 – 601.

[B 72] Sittig, J.: The economic choice of sampling system in acceptance sampling. Bull. Intern. Statist. Inst. 33 (1951) 51 – 84.

[B 73] Smirnov, N. V.: Sur les écarts de la courbe de distribution empirique. Rec. Math. (Matemat. Zbornik) N. S. 6 (1939) 3.

[B 74] Stange, K.: Die wirtschaftliche Verteilung des Prüfaufwands auf die verschiedenen Stufen eines Modells mit mehreren Zufallskomponenten. Studies in Mathematical Statistics. Publishing House of the Hungarian Academy of Sciences, Budapest, 1968.

[B 75] Steinhaus, H.: Quality control by sampling. Coll. Math. 2 (1951) 98—108.

[B 76] Stephens, K. S.—Larson, E. E.: An evaluation of the MIL-STD-105D system of sampling plans. Industrial Quality Control 23 (1967) 310—319.

[B 77] Stevens, W. L.: Control by gauging. Journ. Roy. Statist. Soc. B 10 (1948) 54—108.

[B 78] Störmer, H.: On a test for the exponential distribution. Metrika 5 (1962), 128—137.

[B 79] Tusnády, G.: On the OC function of the acceptance plan designed for a Pólya process of defective items. Studies in Mathematical Statistics. Publishing House of the Hungarian Academy of Sciences, Budapest, 1968. 193—199.

[B 80] Vas, É.: On the efficiency of the sequential probability ratio test. Publ. Math. Inst. Hung. Acad. Sci. 4 (1959) 171—181.

[B 81] Vincze, I.: On Mathematical Statistical Methods of Quality Control of Mass Production. MTA Matematikai és Fizikai Osztályának Közleményei 4 (1954) 513—523 (Hungarian).

[B 82] Vincze, I.: On some questions connected with two-sample tests of Smirnov type. Proceedings of the 5th Berkeley Symposium on Probability and Mathematical Statistics. Vol. 1 (1967) 657—666.

[B 83] Watson, G. S.—Leadbetter, M. R.: Hazard analysis, I. Biometrika 51 (1964) 175—184, II. Sankhya 26 (1964) 101—106.

[B 84] Wetherill, G. B.—Campling, E. E. G.: The decision theory approach to sampling inspection. Journ. Roy. Statist. Soc. B 28 (1965) 381—416.

[B 84a] Wilk, M. B.—Shapiro, S. S.: The joint assessment of normality of several independent samples. Technometrics 10 (1968) 825—839.

[B 85] Žaludova, A. H.: Statistical quality control of production processes using individual sample values. Bulletin de l'Institut International de Statistique 36 (1958) 573—578.

[B 86] Žaludova, A. H.—Režny, Z.—Ullrich, M.: A survey of some recent Czechoslovak work in automatic statistical process control. Journal of Applied Probability 5 (1968) 43—54.

TABLES

[C 1] Bolshev, L. N.—Smirnov, N. V.: Tables of Mathematical Statistics (Russian). Computing Center of the Academy of Sciences of USSR, Moscow, 1968.

[C 2] Dodge, H. F.—Romig, H. C.: Sampling Inspection Tables. 2nd ed. John Wiley and Sons, New York, 1959.

[C 3] Finney, D. J.—Latscha, R.—Bennet, B. N.—Hsu, P.: Tables for Testing Significance in a 2 × 2 Contingency Table. University Press, Cambridge, 1963.

[C 4] Fisher, R. A.—Yates, F.: Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd, London, 1963.

[C 5] Hald, A.: Statistical Tables and Formulas. Wiley, New York, 1960.

[C 6] Molina, E. C.: Poisson's Exponential Binomial Limit. Van Nostrand, New York, 1942.

[C 7] Owen, D. B.: Handbook of Statistical Tables. Addison-Wesley, Reading, 1962.

[C 8] Pearson, E. S.—Hartley, H. D.: Biometrika Tables for Statisticians. University Press Cambridge, 1954.

[C 9] Pearson, K.: Tables of the Incomplete Beta Function. University Press, Cambridge, 1934.

[C 10] Resnikoff, G. J., Liebermann, G. J.: Tables of non-central *t* distribution. Stanford University Press, Stanford, 1957.

[C 11] Romig, H. S.: 50—100 Binomial Tables. Wiley, New York, 1953.

[C 12] Stange, K.—Henning, H. J. (Graf-Henning-Stange): Formeln und Tabellen der mathematischen Statistik. Springer-Verlag, Berlin—New York, 1966.

[C 13] Tables of the Binomial Probability Distribution. National Bureau of Standards, Washington, 1950.

STANDARDS

[D 1] MIL-STD-105D: Sampling Procedures and Tables for Inspection by Attributes. Office of the Assistent Secretary of Defence, Washington, D. C. 1963.

[D 2] MIL-STD-414: Sampling Procedures and Tables for Inspection by Variables for Percent Defective. Office of the Assistent Secretary of Defence, Washington, D. C. 1957.

[D 3] MIL-STD-690A: Failure Rate Sampling Plans and Procedures. Office of the Assistent Secretary of Defence, Washington, D. C.

[D 4] MIL-STD-721: Definition of Terms for Reliability Engineering. Office of the Assistent Secretary of Defence, Washington, D. C.

[D 5] H-106: Multi-level Continuous Sampling Procedures and Tables for Inspection by Attributes. Office of the Assistent Secretary of Defence, Washington, D. C.

[D 6] H-107: Single-level Continuous Sampling Procedures and Tables for Inspection by Attributes. Office of the Assistent Secretary of Defence, Washington, D. C.

[D 7] H-108: Sampling Procedures and Tables for Life and Reliability Testing (Based on Exponential Distribution). Office of the Assistent Secretary of Defence, Washington, D. C.

[D 8] TR-3: Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Mean Life Criterion). Office of the Assistent Secretary of Defence, Washington, D. C. 1961.

[D 9] TR-4: Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Hazard Rate Criterion). Office of the Assistent Secretary of Defence, Washington, D. C. 1962.

INDEX

α risk 290
absolute moment 58
acceptable quality level (AQL) 290
acceptance
  control chart 262
  control limit 262
  number 289
  plans 288
  region 143
  sampling 288
  sampling scheme 293
addition theorem for probabilities 29
adjustment of machine 240
admissible decision function 204
alternative hypothesis 141
AOQ (average outgoing quality) 314
AOQL (average outgoing quality limit) 313
approximation
  normal to binomial 129
  normal to Poisson 277
AQL (acceptable quality level) 290
arithmetic mean 94
  variance of — 67
ASN (average sample number) 305
asymmetry 59
  measure of — 59
attribute, inspection by 263, 289
availability index 373
average
  outgoing quality (AOQ) 314
  outgoing quality limit (AOQL) 313
  sample number (ASN) 305
  sample size 305

β risk 291
Bartlett test 173
Bayes
  method of — 132
  solution 205
  theorem of — for discrete variables 40
    for continuous variables 86
best unbiased estimator 118
beta distribution 81, 137
binomial
  distribution 75, 129
  process 221
  sequence 219
bivariate
  distribution 48
    normal — 195
breakeven quality 315

c (control) chart 275
capability 238
Cauchy distribution 81
cell 90
  frequency 91
  midpoint 91
central limit theorem 73, 85
central line 245
central moments 58
certain event 27
chain sampling 317
characteristic
  operating — (OC) function 151, 290
chart, see control chart
Chebyshev's inequality 68
chi-square distribution 78
  -test 185

choice of intervals between controls 278
class interval, see cell
classical definition of probability 29
coefficient
  correlation — 60, 194
  regression — 199
  of variation 100
coherent structures 375
comparison of distributions (samples) 183, 187
complementary event 19
complete class of decision functions 213
complete system of events 31
composite hypothesis 142
conditional
  distribution 53
  probability 31
confidence
  interval 127
  level 128
consistency of a test 162
consistent estimator 118
consumer's risk 291
continuous
  distribution 44
  materials 276
  random variable 18, 44
  sampling tables 316
  stochastic process 215
control chart 241
  for attributes 263
  for defects 275
  for defects per unit 277
  for demerits 277
  for fraction defective 267
  for individual measurements 248
  for largest and smallest value 255
  for median and range 249
  for midrange 255
  for number of defectives 264
  for number of defects 275
  for order statistics 255
  for standard deviation 255
  for variables 241
  for variance 255
control limits 243
correlation
  analysis 193
  coefficient 60, 194
  function 218
covariance 61
  function 228
Cramér–Rao inequality 122, 209
critical region 143
cumulative sum chart 261
curtosis 59

data
  analysis 233
  diagram 236
  sheet 235, 245, 255
decision 111, 200
  function 200
  space 201
decreasing hazard rate 376
defects
  control chart for number of — 275
degrees of freedom 78
demerit 277
dependent
  events 33, 37
  random variables 48
density
  function 51
  histogram 100
deviation
  mean — 58
  standard — 57
DFR (decreasing failure rate) 376
discrete
  distribution 44
  stochastic process 215
  random variable 17, 44
disjoint events 29
distribution
  Beta — 81
  binomial — 75
  Cauchy — 81
  chi-square — 78
  conditional — 53
  continuous — 44
  discrete — 44
  empirical — 100
  exponential — 77

distribution
    F — 167
    function 45
    gamma — 78
    Gaussian, or normal — 70, 163
    hypergeometric — 84
    Kolmogorov — 47
    logarithmic normal — 82
    marginal — 53
    normal, or Gaussian — 70
    Poisson — 74
    prior — 41
    Student — 80
    t — 80
    uniform — 84
    Weibull — 79
distribution-free test 178
distributions with monotone hazard rate 376
Dodge–Romig sampling plan 313
double
    np chart 270
    sampling 299

efficiency 116
elementary event 47
empirical
    distribution 100
        function 101
    mean 94
    median 96
    variance 98
error
    of the first kind 24, 147
    of the second kind 24, 147
estimation 28, 113
    after selection 372
    interval — 127
    point — 114
estimator 111
    best unbiased — 118
    minimum variance — 117
    unbiased — 115
event
    complementary — 19
    elementary — 47
    independent — 33
    random — 19
expected
    loss 203
    value 66
experiment
    random — 15
exponential distribution 77
exponentiality
    testing for — 182
extreme values
    control chart for — 258

F
    distribution 167
    test 167
factorization property 120
failure
    rate 359
    distribution 359
first kind of error 24, 147
fit
    goodness of — 177
flatness 59
fluctuation
    statistical — 20
formation of inspection lots 312
fraction defective 267
    control chart for — 267
freedom
    degrees of — 78
frequency 19
    histogram 93
function of random variable(s) 62

gamma distribution 78
Gaussian
    distribution 163
    stochastic process 223
goodness of fit 177
grouping
    of data 90
    of sequential test 351
guarantee time 360

hazard
    function 358
    rate 359
histogram 93
    frequency — 93
    density — 100

homogeneity
  test for — 187
Howard count of tomato purée 138
hypergeometric distribution 84
hypothesis 23
  alternative — 141
  composite — 142
  null — 141
  simple — 142
  testing statistical — 140

IFR (increasing failure rate) distribution 376
IFRA (increasing failure rate on the average) distribution 377
impossible event 27
inadmissible decision function 204
incentive effect of acceptance sampling 298
increasing hazard rate 376
  —on the average 377
independence
  of events 33
  of experiments 34
  of random variables 59
inequality of Chebyshev 68
inspection
  by attribute 289
  by variables 317
  level 298
  reduced — 312
  sampling — 288
  tightened — 305
interval estimation 127

k-out-of-*n* structures 376
Kolmogorov's theory 47
Kolmogorov–Smirnov
  distribution 107
  two-sample test 109

laws of large numbers 69
least squares method 198
lemma of Laplace 135
lemma of Neyman and Pearson 153
level
  acceptable quality — 290
  of a confidence interval 128
  probability — 146
  significance — of a test 146
life testing 346, 365
  with replacement 367
likelihood ratio test, see probability ratio test
limit
  confidence — 128
  control — 243
limiting quality (LQ) 290
line
  regression — 197
logarithmic normal distribution 82
loss function 203
lot
  size of — 312
  tolerance proportion defective (LTPD) 290
lower control limit 243
LQ (limiting quality) 290
LTPD (lot tolerance proportion defective) 290

marginal distribution 53
Markov
  chain 216
  process 216
mass phenomenon 16
mean
  arithmetic — 94
  deviation 58, 97
  range 107
mean time
  between failures (MTBF) 359
  to failure (MTTF) 359
median 57
  chart 249
  empirical 96
  theoretical 57
MIL-STD-105-D 293
MIL-STD-414 319
minimal complete class of decision function 213
minimax principle (method) 212
minimum variance unbiased estimator 118
modified control limits 261
moments 58

monotone hazard rate 376
most powerful test 155
MTBF (mean time between failures) 359
MTTF (mean time to failure) 359
multicomponent structures 374
    coherent — — 375
multiple sampling 304
multiplication rule for probabilities 33
multivariate chart 262

narrow-limit gaging 268, 316
natural tolerance 238
Neyman factorization property 120
Neyman–Pearson lemma 153
noncentral t-distribution 170
nonparametric methods 111
normal approximation
    to binomial distribution 129
    to Poisson distribution 277
normal distribution 70
normality
    testing for — 179
normal probability paper 181
np chart 263
    double — 270
null hypothesis 141

observations on a random variable 88
OC (operating characteristic) function 151, 290
operating characteristic curve (function) 151, 290
order statistics 96

p chart 263
parameter 70, 112
    space 70
parametric
    problems 111
    tests 163
percentile 57
periodic review of the control limits 248
phenomena
    random — 15
plan
    sampling — 289
point
    estimation 114
    of control 316

Poisson
    distribution 74
    process 223
Pólya
    process 227
    sequence 228, 355
posterior distribution 41
power function of test 151
preliminary analysis 233
    for attributes 263
    for variables 233
    sampling 23, 233
prior distribution 41
probability 20
    conditional — 31
    level 146, 241
    paper 181
    ratio test 157
process
    analysis 233, 263
        data sheet 235
        data diagram 235
    capability 238
    control
        by attribute 263
        by variables 233
    stochastic — 213
producer's risk 290

quantile 57
quartile 57

random
    event 19
    experiment 15
    numbers 15, 313
        tables of — 380
    phenomena 14
    sampling 15
    selection 15
    variable 16
        continuous — 44
        discrete — 44
        independence of —s 59
    vector 17
randomized test 158
range 99
    distribution of — 105
    in control chart 242

recurrent stochastic process 225
  stationary — 225
reduced inspection 312
regression 193
  coefficient 199
  line 197
  parabola 197
rejected lots 313
rejection
  number 289
  region 143
relative frequency 20
reliability 358
  function 359
  theory 358
renewal
  density 361
  function 361
replacement
  sampling with — 34
  sampling without — 34
revision of control limits 248
risk
  consumer's — 291
  function 203
  producer's — 290
rule of total probabilities 40
runs 260, 261

sample 87
  function 218
  mean 94
  median 96
  size 88
  space 17
  variance 98
sampling
  double — 299
  inspection 288
    by attributes 289
    by variables 317
    Dodge–Romig Table of — 313
    Military Standard Table of — 294, 318
  multiple — 304
  plan 289
  scheme 293
  sequential — 330
sampling
  single — 289
  with replacement 34
  without replacement 34
second kind of error 24
sequential analysis 330
sequential probability ratio test 332
Sheppard correction 98
significance level 146
simple
  hypothesis 142
  loss function 206
  network structure 375
single sampling plan 289
size
  of lot 312
  of sample 88
  of test 146
skewness 59
sloping control lines 255
standard
  deviation 57, 98
  form of normal distribution 72
standardized variable 72
stationary
  recurrent process 225
  stochastic process 217
statistical
  decision functions 200
  fluctuation 20
  quality control, definition 11
stochastic
  connection 194
  convergence 69
  process 213
    binomial — 221
    continuous — 215
    discrete — 215
    Gaussian — 223
    Markov — 216
    Poisson — 223
    Pólya — 227
    recurrent — 225
    stationary — 217
    stationary recurrent — 225
  sequence 215
strongly consistent estimator 124

structure
    function 374
    reliability function 374
sufficient statistics 119
Student (t)
    distribution 80
    test 169
sum of independent random variables 63
system
    hazard rate function — 372
    reliability 372

t (Student)
    distribution 80
    noncentral — 170
    test 169
TBF (time between failures) 359
test
    Bartlett — 173
    χ^2 (chi-square) — 185
    consistent — 162
    distribution-free — 178
    for exponentiality 182
    F — 167
    of goodness of fit 177
    of homogeneity 187
    of independence 188
    of Kolmogorov 178
    of normality 179
    one-sided — 146, 167
    randomized — 158
    uniformly most power fuc — 155
three-sigma control limits 242
tightened inspection 305
time
    between failures 359
to failure 359
total probabilities rule 40
total time on test 346, 366
trial 15
truncated sample 366
truncation of a sequential test 350

unbiased estimator 115
    best — 118
unbiased test 162
uncorrelated random variables 61
uniform distribution 84
uniformly best
    decision function 204
    test 153
upper control limit 243

variable, see random variable
variance 57
    of arithmetic mean 67
    sample — 98
variation coefficient 100
vector
    variable 48
    random — 48

warning limits 259
Weibull distribution 79
Welch test 172

We recommend

MATHEMATISCHE STATISTIK MIT INDUSTRIELLEN ANWENDUNGEN

by
I. VINCZE

In German — 440 pages — 32 figures
13 tables — 17 × 25 cm — Cloth

The book gives an excellent survey of the theory of mathematical statistics, its fundamental principles and its special way of thinking. Furthermore, it provides a storehouse of statistical methods which are inevitably used in solving basic problems both in industrial research and practice, moreover, in many other fields. Examples are worked out in detail in order to show the application of a method, and theoretical problems are well elucidated by examples of their applications in the practice.

DISTRIBUTORS
KULTURA
H-1389 BUDAPEST, P.O.B. 149